Vol. 42, No. 2 March, 1935

PSYCHOLOGICAL REVIEW PUBLICATIONS

Psychological Review

EDITED BY

HERBERT S. LANGFELD, PRINCETON UNIVERSITY

S. W. FERNBERGER, UNIV. OF PENNSYLVANIA (*J. of Exper. Psychol.*)
W. S. HUNTER, CLARK UNIVERSITY (*Psychol. Index*)
JOSEPH PETERSON, GEO. PEABODY COLLEGE (*Psychol. Monog.*)
J. A. McGEOCH, UNIV. OF MISSOURI (*Psychol. Bull.*)

CONTENTS

The Effect of Practice upon Time-Order Errors in the Comparison of Temporal Intervals: HERBERT WOODROW, 127.

The Concept of Reliability: RAYMOND ROYCE WILLOUGHBY, 153.

Perception and Imagination: KATE GORDON, 166.

Expression in Music: A Discussion of Experimental Studies and Theories: KATE HEVNER, 186.

Drive and the Characteristics of Driven Behavior: MERLE HUGH ELLIOTT, 205.

Notes:

A Note on the Recall of the Pleasant vs. the Unpleasant: ELON H. MOORE, 214.

A Note on "The Vectors of Mind": HERMAN A. COPELAND, 216.

PUBLISHED BI-MONTHLY
FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION
BY THE PSYCHOLOGICAL REVIEW COMPANY
PRINCE AND LEMON STS., LANCASTER, PA.
AND PRINCETON, N. J.

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

PUBLICATIONS

OF THE

AMERICAN PSYCHOLOGICAL ASSOCIATION

EDITED BY

S. W. FERNBERGER, University of Pennsylvania (*J. Exper. Psych.*)
WALTER S. HUNTER, Clark University (*Index and Abstracts*)
HENRY T. MOORE, Skidmore College (*J. Abn. and Soc. Psychol.*)
HERBERT S. LANGFELD, Princeton University (*Review*)
JOSEPH PETERSON, George Peabody College (*Monographs*)
JOHN A. McGEOCH, University of Missouri (*Bulletin*)

HERBERT S. LANGFELD, Business Editor

PSYCHOLOGICAL REVIEW

containing original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.

PSYCHOLOGICAL BULLETIN

containing critical reviews of books and articles, psychological news and notes, university notices, and announcements, appears monthly (10 numbers), the annual volume comprising about 720 pages. Special issues of the Bulletin consist of general reviews of recent work in some department of psychology.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

containing original contributions of an experimental character, appears bi-monthly, February, April, June, August, October, and December, the six numbers comprising a volume of about 900 pages.

PSYCHOLOGICAL INDEX

is a compendious bibliography of books, monographs, and articles upon psychological and cognate topics that have appeared during the year. The Index is issued annually in June, and may be subscribed for in connection with the periodicals above, or purchased separately.

PSYCHOLOGICAL ABSTRACTS

appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.

PSYCHOLOGICAL MONOGRAPHS

consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The Monographs appear at irregular intervals and are gathered into volumes of about 500 pages.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

appears quarterly, April, July, October, January, the four numbers comprising a volume of 448 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.

ANNUAL SUBSCRIPTION RATES

Review: $5.50 (Foreign, $5.75). Index: $4.00 per volume.
Journal: $7.00 (Foreign, $7.25). Monographs: $6.00 per volume (Foreign, $6.30).
Bulletin: $6.00 (Foreign, $6.25). Abstracts: $6.00 (Foreign, $6.25).
Abnormal and Social: $5.00 (Foreign, $5.25). Single copies $1.50.
Current numbers: Journal, $1.25; Review, $1.00; Abstracts, 75c; Bulletin, 60c.

COMBINATION RATES

Review and Bulletin: $10.00 (Foreign, $10.50).
Review and J. Exp.: $11.00 (Foreign, $11.50).
Bulletin and J. Exp.: $12.00 (Foreign, $12.50).
Review, Bulletin, and J. Exp.: $16.00 (Foreign, $16.75).
Review, Bulletin, J. Exp., and Index: $19.00 (Foreign, $19.75).

Subscriptions, orders, and business communications should be sent to the

PSYCHOLOGICAL REVIEW COMPANY

PRINCETON, NEW JERSEY

VOL. 42, NO. 2 MARCH, 1935

THE PSYCHOLOGICAL REVIEW

THE EFFECT OF PRACTICE UPON TIME-ORDER ERRORS IN THE COMPARISON OF TEMPORAL INTERVALS

BY HERBERT WOODROW

University of Illinois

Recent studies of time-order errors show that the significance of the peculiar and striking behavior of these errors for the theory of comparison is fully recognized. Apparently it is not so fully recognized that in no field are these phenomena more remarkable or varied, and in no field have they been more intensively investigated, than in the case of temporal perception. For example, the increase in the time-order error with increase in the length of pause between the first and the second of two compared objects, a phenomenon described by Köhler in 1923 [1] and extensively used by him in arguing in favor of his widely known theory of the successive comparison of intensities, appears to have been demonstrated with fair dependability by Katz [2] as far back as 1906. It seems quite possible that had Köhler taken into consideration the results obtained by Katz, which show originally positive errors becoming increasingly *positive* with increase in the pause, he would not have offered a physiological theory which could readily account only for an increase in the *negative* errors with increase in pause.

On the other hand, the effect of practice upon time-order

[1] W. Köhler, zur Theorie des Sukzessivvergleichs und der Zeitfehler, *Psychol. Forsch.*, 1923, 4, 115-175.

[2] D. Katz, Experimentelle Beiträge zur Psychologie des Vergleichs im Gebiete des Zeitsinns, *Zsch. f. Psychol. u. Physiol. d. Sinnes.*, 1906, 42, 302-340 and 414-450.

errors has already been studied, notably by Köhler,[3] in the case of sound intensities, but seems not hitherto to have been observed in the case of temporal attributes. The data to be presented, however, show in the case of temporal discrimination a notable shift in time-order errors with practice, and while these data are limited to a single pause of 2 seconds, they give in some respects more complete information concerning this shift than has heretofore appeared, in that they show the shift separately for each variable, including variables longer as well as shorter than the standard, and show the extent to which the shift occurring at one sitting carries over to a later sitting. In the case of temporal comparisons, data on this phenomenon have a special significance, in that they demonstrate that the value obtained for the temporal indifference interval—the object of so many investigations—depends largely upon the length of the sittings given the subjects. An enormous change may occur even between the first and last part of a sitting which occupies in its entirety only 45 minutes, so that the indifference interval when calculated from data based on sittings, say, of 30 trials with a given standard, may differ by several hundred sigma from the interval calculated from sittings of 120 trials. After an extremely thorough investigation of the indifference interval in the case of tonal durations, Stott[4] concluded that the effects of practice were largely responsible for a shift in the indifference interval from about 1500 σ in the case of highly practiced subjects to about 900σ in the case of short sittings with naive subjects.

Practice also has a very peculiar effect upon the difference limens. In the data presented below, in spite of practice sufficient to cause in general an increase in acuity of approximately 29 per cent, two limens, namely, the upper difference limen with the variable 2d, and the lower difference limen with the variable 1st, were larger at the end of this practice than at the beginning of it!

These striking phenomena appear to fit in perfectly with

[3] *Op. cit.*, 159–165.

[4] L. H. Stott, Time-order errors in the discrimination of tonal durations, *J. Exper. Psychol.*, in press.

certain general postulates of a theory of the successive comparison of intensities already advanced by the writer,[5] and since a satisfactory explanation of the phenomena does not seem to have been previously given, an explanation in terms of these postulates is here developed.

The present data were obtained in the course of an extensive experiment on the transference of training with reference to factorial analysis, in which 110 subjects, university undergraduates, were given a long series of group tests, then divided into four practice sections, and after being practiced in one or more tests, were again all given a long list of tests. The 22d of the list of tests consisted of 120 comparisons of empty intervals, instituted by the method of constant stimuli, with a standard of 1000 σ. This was the first test of a temporal nature in the series. At the conclusion of the first long series of tests, 28 of the 110 subjects were given practice in both pitch discrimination and time discrimination. The purpose of this practice is of no concern in the present connection, and it is not believed that the practice in pitch discrimination could by any possibility be responsible for the phenomena, observed repeatedly in the time-discrimination sittings, demonstrated by the data to be here presented.

The data obtained with the 110 subjects at the first sitting illustrate clearly enough the phenomenon of the shift in the time-order errors, but data obtained with the practice group will also be given, since the latter data not only repeatedly verify the existence of the phenomenon revealed in the first sitting, but add additional information as to the effect of practice thereon.

The time-intervals used in this study were empty intervals bounded by the instantaneous sounds produced by the arm of an electromagnetic relay. The interval between any pair of sounds, as well as the pause between the first and second of any two compared intervals, was controlled by means of the spacing of contacts attached to the rim of a Meumann's time-sense apparatus, the arms of which were rotated by means of a

[5] Weight-discrimination with a varying standard, *Amer. J. Psychol.*, 1933, 45, p. 409.

synchronous motor. The pause between the first and second intervals of any comparison pair was 2 seconds. At each trial there was given 2 seconds before the beginning of the first interval, a ready signal, consisting of the momentary flash of a miniature light bulb suspended on the black-board in front of the group. The loudness of the sounds was the same as that used in a previous study [6] of the temporal indifference interval by the method of reproduction, or mean error; though, since the subjects were tested in groups of 55 in the case of the data represented in Table I and in a group of 28 in the remaining cases, the loudness must have varied considerably with the position of the individuals in the room.

With a standard of 1000 σ, there were used five variables, differing from the standard in percentage of length by -20, -10, 0, $+10$ and $+20$.

The data obtained with the group of 110 subjects, given in Table I, represent the 13,200 judgments obtained in a series of 120 trials, run off without intermission. The series consisted, however, of a fixed list of thirty pairs of intervals given four times over. In this set of 30 trials, each of the five variables was given six times, three times before the standard and three times following it.

The judgments were first tabulated in terms of three categories, and then reduced to two, 'longer' and 'shorter' by dividing the 'equal' judgments evenly between these two categories. The data will be presented, however, so as to show for each variable how often the judgments signified that the variable was longer. When the variable is given first, the percentage of judgments signifying that it is the longer is of course the percentage of judgments of the category 'second shorter'. The difference in the case of the two orders, V 1st and V 2d, in the percentage of judgments meaning that the variable seemed longer, affords an index of the extent of the time-order error. Each percentage is based on 330 judgments, 3 from each of 110 subjects, except the percentages given for the variable differing by zero from the standard. In the case

[6] The temporal indifference interval determined by the method of mean error, *J. Exper. Psychol.*, 1934, 17, 167–188.

of zero difference, the labelling of one of the intervals as the standard and the other as the variable, is of course entirely arbitrary; consequently all trials with zero difference, whether labelled standard first or standard second, were pooled into one group of 660 judgments. The percentage of these 660 trials, in which the standard seemed longer when first, is given in the column headed V 1st and the percentage in which it seemed longer when second in the column headed V 2d, so that, as in the case of all the other variables, the time-order error is obtained with its proper sign by subtracting the percentage in the column V 2d from that in the column V 1st. The absolute magnitude of the errors should not, however, be assumed to vary in direct proportion to the differences in the percentage of variable longer judgments; for when the percentage of variable longer judgments approaches either zero or one hundred, a given difference between the two orders in this percentage indicates a relatively greater time-order effect, in terms of sigma, than when the percentage is near fifty. In labelling these differences, given in the columns headed 'Error', plus and minus, traditional procedure has been followed: so that a negative error means that the judgments signified that the variable seemed longer when given after the standard and a positive error that it seemed longer when given first.

The data of Table I show that during the sitting there was a somewhat gradual and irregular, but very pronounced, shift in the time-order errors from the negative direction towards the positive one. The average of the errors for all five variables decreased from −15.7 per cent in the first 30 trials and −7.7 per cent in the second 30, to −3.8 and −4.7 per cent in the third and fourth 30 trials. Most interesting is the change in the sign of the error in the case of the variables shorter than the standard. For the first 30 trials, both of these variables showed negative time-order errors; for the second 30, only the longer of them, the −10 per cent variable; while for the third and fourth sets of trials they both showed positive time-order errors. In general, then, the error in the case of the shorter variables changed from negative to positive, while the

TABLE I

CHANGE IN TIME-ORDER ERRORS DURING THE FIRST SITTING WITH 110 SUBJECTS

Period	Variable	V 1st	V 2d	Error
1st 30 trials	800	11.8	17.9	− 6.1
	900	27.4	38.8	−11.4
	1000	40.6	59.4	−18.8
	1100	53.9	74.5	−20.6
	1200	68.2	90.0	−21.8
2d 30 trials	800	14.8	11.1	+ 3.7
	900	32.7	33.8	− 1.1
	1000	45.9	54.1	− 8.2
	1100	59.7	75.1	−15.4
	1200	68.6	85.9	−17.3
3d 30 trials	800	17.6	5.5	+12.1
	900	38.9	30.3	+ 8.6
	1000	43.8	56.2	−12.4
	1100	58.8	72.4	−13.6
	1200	71.4	85.0	−13.6
4th 30 trials	800	17.0	9.7	+ 7.3
	900	32.9	30.8	+ 2.1
	1000	48.2	51.8	− 3.6
	1100	55.8	69.4	−13.6
	1200	70.6	86.1	−15.5

error with the longer variables and with zero difference became less negative.

The data of Table II show what happens at repeated sittings. Since the additional sittings were given to only 28 of the subjects, the data for these particular subjects at the first sitting are included in the Table. The results for the 28 subjects at the first sitting agree roughly with those for the entire group of 110 subjects, and show the same general trends, including the reversal of the sign of the error in the case of the shorter variables with prolongation of the sitting. In Table II, the first sitting is called sitting *A*, and later sittings *B*, *C* and *D*. Various amounts of practice intervened between the sittings, some of it corrected practice.

1. (*A*) April 10. Initial sitting, any sort of temporal discrimination.
2. April 11. Seashore Time and Rhythm tests.
3. April 12. 100 trials, filled intervals, standard 850 σ.
4. April 17. 25 trials, empty intervals.
5. April 18. 12 trials, empty intervals.
6. (*B*) April 19. 90 trials (results included in Table II).
7. April 20. 30 trials. The same 30 trials then repeated, with subjects looking at the record of their original reports, and the correct report announced immediately after each trial.

TABLE II

CHANGES IN TIME-ORDER ERRORS WITH PRACTICE AT THE SAME SITTING AND ON SUCCESSIVE SITTINGS

Sitting	Period	Variable	V 1st	V 2nd	Error
A (120 trials)	1st 30 trials	800	19.6	20.8	− 1.1
		900	26.8	34.5	− 7.7
		1000	40.0	60.0	−20.0
		1100	46.4	76.2	−29.8
		1200	62.5	85.7	−23.2
	2d 30 trials	800	16.7	15.5	+ 1.2
		900	28.0	40.5	−12.5
		1000	50.2	49.8	+ 0.4
		1100	67.2	76.2	− 9.0
		1200	68.7	87.5	−18.8
	3d 30 trials	800	23.2	7.1	+16.1
		900	39.3	29.2	+10.1
		1000	44.0	56.0	−12.0
		1100	60.1	68.4	− 8.3
		1200	72.3	81.5	− 9.2
	4th 30 trials	800	20.8	10.1	+10.7
		900	30.9	30.3	+ 0.6
		1000	46.9	53.1	− 6.2
		1100	59.5	67.2	− 7.7
		1200	65.2	85.1	−19.9
B (90 trials)	1st 30 trials	800	10.2	8.3	+ 1.9
		900	28.8	30.7	− 1.9
		1000	49.5	50.5	− 1.0
		1100	58.2	78.1	−19.9
		1200	66.3	84.5	−18.2
	2d 30 trials	800	17.9	8.3	+ 9.6
		900	28.8	30.1	− 1.3
		1000	48.3	51.7	− 3.4
		1100	62.7	69.7	− 7.0
		1200	74.2	80.0	− 5.8
	3d 30 trials	800	19.8	8.3	+11.5
		900	40.3	21.1	+19.2
		1000	50.0	50.0	0.0
		1100	62.7	73.6	−10.9
		1200	71.7	79.4	− 7.7

8. April 23. 30 trials, repeated, with correct answers announced the second time. 81 more trials following this corrected practice.

9. (*C*) April 24. 120 trials, with variables differing from standard by 8 and 16 per cent instead of 10 and 20 per cent as heretofore.

10. (*D*) April 26. 30 trials, with same series as on April 24.

TABLE II (*cont.*)

Sitting	Period	Variable	V 1st	V 2nd	Error
C (120 trials)	1st 30 trials	840	37.6	37.0	+ 0.6
		920	34.0	46.3	−12.3
		1000	49.7	50.3	− 0.6
		1080	56.2	69.1	−12.9
		1160	64.8	75.3	−10.5
	2d 30 trials	840	29.8	12.5	+17.3
		920	37.5	20.8	+16.7
		1000	45.2	54.8	− 9.6
		1080	59.5	69.6	−10.1
		1160	70.8	72.6	− 1.8
	3d 30 trials	840	23.8	6.5	+17.3
		920	40.5	23.2	+17.3
		1000	48.8	51.2	− 2.4
		1080	51.8	61.9	−10.1
		1160	78.5	89.3	−10.8
	4th 30 trials	840	33.3	4.8	+28.5
		920	39.9	20.8	+19.1
		1000	48.2	51.8	− 3.6
		1080	62.5	58.9	+ 3.6
		1160	75.0	83.9	− 8.9
D (30 trials)	30 trials only	840	16.1	4.2	+11.9
		920	31.5	32.7	− 1.2
		1000	45.8	54.2	− 8.4
		1080	62.5	79.1	−16.6
		1160	75.0	83.9	− 8.9

The general trend of the errors shown in Table II may be readily observed in Table III, in which the time-order errors

TABLE III

CHANGE IN AVERAGE TIME-ORDER ERROR WITH SITTING AND WITH DURATION OF SITTING

	1st 30	2d 30	3d 30	4th 30
Sitting *A*	−16.4	−7.7	−0.7	−4.5
Sitting *B*	− 7.8	−1.6	+2.4	
Sitting *C*	− 7.1	+2.5	+2.8	+7.7
Sitting *D*	− 4.6			

of all variables for each 30 trials have been averaged. It will be noted that in each row, reading from left to right, the errors shift so as to become less negative or more positive, and likewise that in each column, reading down, the errors shift in this same direction. There is only one exception to the regularity

of these changes, namely, an increase in the size of the negative error from the 3rd to the 4th 30 trials of sitting *A*. In general, then, at every sitting the errors shifted in the positive direction with prolongation of the sitting and at each successive sitting they also became more positive (or less negative) than at the preceding sitting.

Between sittings, a considerable part of the effect of the previous sitting or sittings was lost, as is clear from the fact that errors for the initial 30 trials of any sitting were more negative than the error for the final trials of the preceding sitting. On the other hand, a considerable part of the effect of the previous sittings was retained, as is demonstrated by the fact that the errors for the initial 30 trials, as likewise those for the 2d, 3rd or 4th 30 trials, of any sitting were less negative, or more positive, than those for the corresponding period of the previous sitting.

It should be noted, however, that in spite of the shifting of the errors for all variables in the positive direction, the only variables which ever came reliably to show errors of positive sign were the two variables shorter than the standard. In the case of these variables the errors changed from an average of -4.4 per cent in the first 30 trials of the first sitting to $+5.7$ per cent in the last 30 trials of that sitting, to $+15.4$ per cent in the last 30 trials of sitting *B*, and to $+23.8$ per cent in the last 30 trials of sitting *C*. For the variables longer or equal to the standard, the change consisted simply in a decrease in the size of the negative error, *e.g.*, in the case of the two variables longer than the standard, from an average of -26.5 per cent in the first 30 trials of sitting A to one of -2.7 per cent in the case of the last 30 trials of sitting C. The zero difference gave rather irregular results, but with the exception of the last 30 trials of sitting B, where the error was zero, continues throughout to show a negative error.

While the phenomenon which has been described above is a result of what might be termed practice, it cannot be explained as consisting of, or resulting from, what ordinarily would be considered the effect of practice, namely, an increase in acuity. It is of course true that acuity is ordinarily better with the

variable second than with the variable first. A greater acuity with variable 2d would of itself produce negative time-order errors in the case of variables longer than the standard and positive errors in the case of variables shorter than the standard and zero error when there was no difference between standard and variable. This distribution of errors was not, however, the one found at the beginning of the present experiment. To begin with, all variables gave negative errors; and the shift from negativity towards positivity with prolongation of the trials, which constitutes the phenomenon under discussion, could be explained as an ordinary practice effect only providing practice improved acuity in the case of V 2d much more decidedly than in the case of V 1st. Moreover, since the shift in the time-order error manifests itself practically throughout the whole experiment, relatively greater improvement with V 2d would have to be present with great regularity and manifest itself at any two stages of the experiment which might be chosen for calculating the degree of improvement. As a matter of fact, however, the shift phenomenon manifests itself in marked degree both in the absence of any increase in acuity in the case of either V 2d or V 1st and in cases where the improvement with V 1st is very nearly equal to that with V 2d. For example, in the case of the sitting with 110 subjects (Table I), there is no improvement in acuity (measured by the per cent of correct judgments in the case of variables other than the variable of zero difference) between the first half and the second half of the sitting. The longer variables gave 71.9 per cent correct judgments in the 1st half and 71.2 per cent correct judgments in the 2d half, while the shorter variables gave 76.5 per cent correct in the first half and 77.2 per cent correct in the second half. In short, there was no appreciable improvement in the case of either V 1st or V 2d; and yet the time-order error for the shorter variables in the first half was −3.7 per cent and in the second half +7.5 per cent.

Again, in the case of the 28 subjects given a number of sittings, the improvement between the first 60 trials of sitting *A* and the last 60 trials of sitting *B* consisted in an increase in the percentage right with V 1st of 1.4 per cent (from 69.2 to

70.6), and with V 2d of 2.6 per cent (from 76.8 to 79.4). The difference in degree of improvement for the two orders, V 1st and V 2d is practically negligible, and cannot possibly account for the shift in the average time-order error from −12.1 per cent to +0.4 per cent.

In general, then, the phenomenon in question is not due to gain in acuity with V 2d beyond the gain with V 1st. The effect of practice between certain points does show such relative gain for V 2d over V 1st, but the shift phenomenon is present in very striking fashion when no such relatively greater improvement with V 2d is to be noted. It cannot therefore be explained as due to a change in the relative acuity with the two orders V 2d and V 1st, but must be regarded as occurring independently of any such change. While as a matter of convenience in terminology, the shift phenomenon will hereafter be referred to occasionally as a practice effect, it should be understood that the word practice is used merely in the sense of prolongation of the series of trials.

The true explanation of the shift in time-order error requires reference to certain aspects of the process of successive comparison. In the comparison of both intensities and temporal magnitudes, the subject's judgment depends by no means solely upon the difference between the two stimuli (or perceptual objects) compared. It depends rather upon the relation of the second magnitude to the magnitude the first would have if it were supposed to have gravitated during the pause between the first and the second towards one or more other magnitudes, these latter magnitudes consisting of what may be termed remote or background standards. To go into the explanation of this gravitation of the magnitude given first, during the pause between the first and second stimulus-object, would require a rather elaborate discussion of the physiological theory both of comparison and of the perception of empty intervals. It is unnecessary here to enter such discussion. It is possible to offer an explanation, of a schematic or 'as though' type, without going into more ultimate theory; and there is the advantage in such an explanation that it will adhere closely to established facts.

The chief fact which must be kept in mind is the one already mentioned that the percentages of the various categories of judgments change with the order of presentation of two stimuli as though the value of the first stimulus changed during the pause between it and the second stimulus. The value which the first magnitude appears to have acquired at the moment that the second one is presented may be termed the effective standard. Commonly, at least with pauses longer than 3 seconds, the first magnitude will appear to have dropped to some lower level, or, differently expressed, to have gravitated towards some lower remote standard, so that the effective standard is of lesser magnitude than the first stimulus.

In the present instance, it is plausible to assume that this lower remote standard is the absolute indifference interval, a remote standard considerably lower than the experimental standard of 1000 σ. This assumption is plausible because an indifference interval is by definition precisely an interval towards which the first interval of a comparison pair [7] appears to gravitate—upwards when the first interval is shorter than the indifference interval and downwards when the first interval is longer. In addition, however, there is another remote standard consisting of the average magnitude of the entire experimental series. Experimental evidence establishing the unavoidability of this hypothesis of the series average as a remote standard has been published both by Lauenstein [8] and the writer.[9] This series average length, which in the

[7] In the case of 'absolute' judgments, it may be assumed that the effective standard with which each stimulus is compared is the series average. In case there were present also an absolute indifference magnitude, the series average would, during the pause between stimuli, gravitate towards this magnitude. Even in the present experiment, it is a matter of indifference whether the series average or the first interval is regarded as gravitating towards the absolute indifference interval, since the first interval is itself a part of the series. As stated in a previous discussion, "The average of the entire series should be thought of as a weighted average, in determining which, the most recent stimuli have the greater weight and the last stimulus [before the second member of a comparison pair] the greatest weight"—H. Woodrow, Weight discrimination with a varying standard, *Amer. J. Psychol.*, 1933, 45, p. 409.

[8] O. Lauenstein, Ansatz zu einer physiologischen Theorie des Vergleichs und der Zeitfehler, *Psychol. Forsch.*, 1933, 17, p. 174.

[9] *Op. cit.*, 404–415.

present work was exactly equal to the length of the standard, was given especial emphasis in that it occurred in the stimulus series, as the length of the standard, six times as often as any of the four variables longer or shorter than the standard (since the standard was given first in 50 per cent of the trials and second, as the zero difference variable, in 10 per cent of the trials).

In order to explain the shift in the time-order error with increase in the number of trials, it is necessary only to assume that as the subject becomes habituated or adapted to the average length of the experimental series, such habituation weakens the initially relatively powerful influence of the absolute indifference interval as a remote standard. The series average could not initially be as powerful a remote standard as the absolute indifference interval, since the series average is set up as a remote standard only through the repeated presentation of the intervals composing that series. Initially, then, the absolute indifference interval would exert a more powerful gravitational influence than the series average. But as the subject became adapted to the length of the experimental standard, or series average, this new adaptation would gradually and, in the case of any individual subject, more or less irregularly and imperfectly replace the original and customary adaptation to the length of the indifference interval.

It is a simple matter to show how these assumptions would lead to the time-order errors and the changes therein found in the present experiment. Gravitation towards the series average would cause the longer variables to become shorter, the shorter ones to become longer, and leave the zero difference variable unchanged. This factor taken alone, then, would result in a negative time-order error for the longer variables, zero error for the zero difference variable, and a positive time-order error for the shorter variables.

The influence of the standard, however, is relatively slight at the beginning of a sitting, while the effect of the absolute indifference interval is powerful. The latter alone, since it is shorter than any of the variables, would cause a large negative

time-order error in all variables. The error resulting from the combined influence of the two remote standards, therefore, is a combination of the large negative error in the case of all variables, due to gravitation towards the absolute indifference interval, with errors of varying direction due to gravitation towards the series average. In the case of longer variables, both remote standards tend to produce negative time-order errors; therefore the effect of one enhances that of the other, and the resultant errors are very decidedly negative. In the case of the zero difference variable, one remote standard, the series average, tends to produce zero error and the other, the absolute indifference interval, a negative error, and the result is a negative error. In the case of the shorter variables, the series average and the absolute indifference interval exert contrary effects, but since at the beginning of the sitting the absolute indifference interval has more effect than the standard the resultant is a negative error, but a much smaller negative error than in the case of the longer variables. With prolongation of the experiment the effect of the absolute indifference interval is largely blotted out as the result of habituation to the standard. As a consequence, towards the end of a long sitting, all errors are less negative than at the beginning, and those with the shorter variables, due now primarily to gravitation upwards toward the standard, become positive.

Strong evidence in support of this explanation of the shift in errors with increase in the number of trials is afforded by the results obtained by Stott,[10] who studied this matter in the case of the comparison of continuous tonal durations at three different sittings. In one of these,[11] the initial trials showed an indifference interval near the middle of the series of variables, that is, near the series average. Even in the first 10 trials, the two shorter variables gave positive time-order errors and the three longer ones negative time-order errors. Obviously, in such a case, where the absolute indifference interval and the series average very nearly coincide, practice should not cause any great change in the errors; for when the two remote

[10] *Op. cit., J. Exper. Psychol.*, in press.

[11] The case described under the heading, Experiment IV.

standards, the series average and the absolute indifference interval coincide, the direction of the errors produced by each of the remote standards would coincide. Therefore, if, as assumed above, it is the change in the relative influence of these two remote standards with prolonged sittings which accounts for the shift in the time-order errors, such change would leave the errors unaffected. It should be mentioned that Stott in this case used an asymmetrical series of variables, all the variables being longer than the standard; but this fact does not explain the absence of a practice effect upon the time-order errors. Köhler likewise used an asymmetrical series of variables, without thereby destroying a marked effect with practice upon the time-order errors.

In his other two cases, in one of which the indifference interval lay far below the series average and in the other of which it coincided with the shortest variable (a variable 20 per cent shorter than the standard), Stott found a pronounced shift in the time-order errors quite analogous to the one described in the present paper.

It should be emphasized again that in the explanation offered above the term gravitation is not used to indicate any hypothetical, physiological causative agency, but simply as meaning that the judgments turn out as though the interval presented first had, during the pause between the two compared intervals, grown either longer or shorter. And when it is said, *e.g.*, that the first gravitates towards the indifference interval, all that is meant is that the explanation of why the judgments turn out *as though* the first had become shorter, if fully stated, would have to take into account the existence of this indifference interval. Obviously numerous other words or phrases could be substituted for the word gravitation. Among these are assimilation, sinking or rising towards, or the concept of the partial or occasional substitution of one of the so-called remote standards for the actually given first interval. Moreover, no implication is here intended concerning what actually persists in the subject during the pause, and which makes possible a comparison with the second interval. Whether this be a 'trace', an accumulation of 'sediment', or an

attunement or adjustment, it is something that changes during the pause between the first and second interval, and inasmuch as the exact physiological nature of the change is unknown, it is simply a verbal convenience to say that the judgments turn out as though the first interval had changed.

The present data are, as previously indicated, analogous to those of Köhler, who found that, in the case of comparison of sound intensities, at successive sittings the time-order errors became more positive or less negative. Köhler's data show that, whereas in the initial trials the errors were positive only with pauses of less than 3 seconds between the first and second stimulus, with practice they became positive with pauses as long as $4\frac{1}{2}$ seconds, while with a still longer pause, the errors, though not changing to positive, nevertheless changed to much less pronounced negative ones. Köhler took the view that the explanation of the trend of the errors towards the positive direction with practice should be due to an extension of the range of action of the same cause as that producing the initially present positive error with intervals under 3 seconds. This view is of no great explanatory value, however, since Köhler was unable to decide upon the cause of the initially positive errors. Lauenstein, in fact, in his recent contribution from Köhler's laboratory, states that the explanation of the positive error with short pauses remains today completely obscure—'volkommen unklar.'[12] The practice effect noted by Köhler is very readily explained, however, by the concept of gravitation towards the series average.

In Köhler's series all the variables were of lower magnitude than the standard. In such a series, gravitation towards the series average, should, when the results for all variables are pooled, as by Köhler, cause a positive error. If then, as has here been alleged, gravitation towards the series average becomes with practice more exclusively the sole cause of the errors, the errors would with practice become more positive.

Whether the explanation of the effect of practice here offered will also serve as an explanation of the positive time-order error which Köhler found at the initial trials with pauses

[12] *Op. cit.*, p. 169.

under 3 seconds cannot with certainty be determined. If the phenomenon were as Köhler apparently conceives it to be, that is, one which manifests itself irrespective of whether the variables are longer or shorter than the standard, then it would not. For gravitation towards the standard, with an interval long enough to allow any appreciable effect, should cause *negative* errors in the case of longer variables and positive errors only in the case of shorter variables. Several recent writers apparently conceive the phenomenon of positive time-order errors with intervals under 3 seconds to be a general one, observable, that is, in all cases of successive comparison of magnitudes. Such writers have overlooked the field of temporal comparisons, in which it is clear that no such generality attaches to the phenomenon in question. The present data, for example, reveal a definitely *negative* initial time-order error with an interval as short as 2 seconds. It is true that the shorter variables display a smaller negative error than the longer variables, and so may be considered to have been affected by a factor *tending* to produce positive errors; but there is no evidence even of a *tendency* towards positive errors in the case of the longer variables. Köhler, however, used an asymmetrical series in which the variables were of lower intensity than the standard; consequently it is possible that his positive errors with very short pauses were due to gravitation of the first intensity towards the standard during the pause. Such an hypothesis might even explain the fact that the positive time-order errors did not reach their maximum at once but only after a pause of 1½ seconds. Before such an explanation should be advanced, however, further information should be obtained concerning the effect of the length of pause upon the relative strength of influence of the series average and a lower remote standard. It seems rather plausible that any absolute standard would gain in influence relative to the series average with increase in the length of the pause, that is, with increase in the interval since the last stimulation, since the series average as a remote standard is a temporary unstable level of reference, not only produced by the stimulation composing the series, but largely dependent

for its continuance upon the continuance of these stimulations. Hypotheses on this matter, however, are hardly justified before further investigation. But whatever may be the explanation of the positive errors found by Köhler with very short intervals, the shift in errors which he observed with practice (continuation of the sittings) appears to be much more definitely explained by the hypotheses here adopted than in any other manner hitherto suggested.

These same hypotheses account for the negative time-order errors found by Köhler and others with pauses longer than 3 seconds. Now Köhler ascribes the negative time-order error to a sinking of 'silent traces,' and does not use any such concept as gravitation towards a remote standard. And, indeed, it may seem rather far-fetched even in an explanation such as here offered, an explanation that aims simply to bring particular phenomena into line with a generalized description of a broader set of tendencies, without attempting any more fundamental physiological explanation of these general tendencies, to describe a shift in the effective standard as a shift towards a low, remote standard, when no such standard has been definitely established. But the fact that no such remote standard has been proved to exist in the case of tonal intensities does not alter the fact that the judgments turn out as though the first stimulus did shift toward such a hypothetical standard. Moreover, Lauenstein, who presents a theory of which Köhler himself apparently approves, definitely comes to the conclusion that apparent absence of stimulation acts as a remote standard, which, in different terminology than here used, he calls 'die Spur des Zustandes der Reizlosigkeit,' and he regards the first stimulus as undergoing a process of approaching (Annäherung) or assimilation (Angleichung) with respect to this hypothetical standard. It appears possible that there may exist a remote standard even more analogous to the temporal indifference interval than the one pointed out by Lauenstein. Katz,[13] in 1906, came to the conclusion that such absolute indifference levels existed in the case of all modes

[13] D. Katz, Experimentelle Beiträge zur Psychologie des Vergleichs im Gebiete des Zeitsinns., *Zsch. f. Psychol. u. Physiol. der Sinnes.*, 1906, 42, p. 436.

of sensory experience. He regarded Wreschner's results, as well as certain series of trials described by Martin and Müller, as experimentally establishing the existence of an indifference level in the case of lifted weights. More recently Pratt [14] has published data interpreted by that author as demonstrating an absolute indifference standard in the case of tonal intensities. The whole subject, however, needs further investigation before one is justified in drawing final conclusions.[15] Should there be cases, however, in which Katz's contention, that indifference values exist in all sensory fields, is not confirmed, then it may be necessary to fall back in those cases upon some such hypothesis as that of Lauenstein, that is, gravitation towards the level represented by a condition of apparent absence of stimulation.

The question of the origin of indifference levels is a very old one. Pratt cites the writer's experiments on weight discrimination as indicating that they are built up during the experiment proper. But the different remote standards should not be confused. The level built up during the experiment is that represented by the series average. This may, it is true, exert an effect lasting for some time afterwards, and, as shown by the present data, influence sittings on following days. Such a temporary remote standard or level should, however, be distinguished from the more permanent and habitual levels which are designated as absolute indifference levels. These latter may be built up in past experience, as Katz [16] and others before him, notably Wundt, have held to be the case. The writer would be inclined to go further and, while conceding the effect of past experiences, regard the absolute indifference levels as largely determined by innate constitution, not simply of the sense-organs, but of the entire sensory-cerebral-motor

[14] Time-errors in the method of single stimuli, *J. Exper. Psychol.*, 1933, 16, 798–814.

[15] Pratt's interesting investigation should be repeated in such a way as to eliminate any possible effect of trials at one level of intensity on the trials with the same subjects at a different level, an effect which might conceivably lead to a quite different interpretation of the results than the one offered by this investigator.

[16] "Unsere Sinnesorgane haben sich im Laufe der Erfahrung an Reize von einer gewissen Stärke gewohnt, die man als mittlere zu bezeichnen pflegt"—Katz, *op. cit.*, p. 436.

mechanism. In the case of time, at any rate, it is hard to regard the indifference interval as produced altogether by experience, since the indifference interval for filled intervals appears to be about 300 σ longer than that for empty intervals, the former being about 900 σ [17] and the latter, at least when determined by the method of mean error, about 625 σ.[18] If, with Wundt, we explain the temporal indifference interval as built up through the constantly repeated experiences of the time required for the swing of the leg, in walking, which is said to average about .5 second, why is it that another indifference interval crops out in the case of the comparison of tonal durations? To the writer it seems more likely that the nervous mechanism is so constructed that it responds twice in succession to momentary stimuli more readily or more adequately when these follow each other, in the case of instantaneous sounds, at an interval of about .6 second than at other rates, whereas it is best adapted in the case of continuous tonal durations to a stimulation lasting in the neighborhood of .9 second. With practice, adaptation to other conditions could be more or less perfectly achieved, but there would be constantly the tendency to revert to the natural attunement due to innate physiological conditions. Such a view, further, accords well with the theory of the general cause of time-order errors advocated by the writer, [19] a theory which explains these errors in terms of the magnitude of stimulus to which the subject is attuned,[20] and to shifts in this attunement produced by experience.

Aside from the effect of practice on time-order errors there are one or two other points worthy of mention. Of some interest is the fact that the judgments signifying the variable to be longer are not distributed normally in accordance with the phi-gamma hypothesis, but, as found by the writer to be

[17] L. H. Stott, *op. cit.*

[18] H. Woodrow, The temporal indifference interval determined by the method of mean error, *J. Exper. Psychol.*, 1934, 17, 167–188.

[19] Weight-discrimination with a varying standard, *Amer. J. Psychol.*, 1933, 45, 408 *ff.*

[20] This view is also in accord with the theory of Katz, who refers to indifference levels as phenomena of *Einstellung*. *Op. cit.*, p. 438.

the case in several previous studies [21] of temporal judgments, are negatively skewed, that is, the median lies lower in the scale of temporal magnitudes than the mean. The results for the entire first sitting with 110 subjects (data for V 1st and V 2d amalgamated) show that the 20 per cent shorter variable was indicated to be shorter (including half the equal judgments) correctly in 86.8 per cent of all the trials, whereas the 20 per cent longer variable was correctly indicated in only 78.3 per cent. Even in the case of the data for the order V 2d taken separately, in the last 30 trials of the sitting the −20 per cent variable was indicated correctly in 90.3 per cent of the trials whereas the +20 per cent variable was correctly indicated in only 86.1 per cent. This greater acuity for decreases than for increases in length shows how hazardous is the procedure, which has sometimes been employed in determining the indifference interval, of taking a point half-way between the upper and lower just noticeably different intervals as the subjectively equal interval and then basing conclusions concerning the time-order error solely on the difference between the subjectively equal standard as thus determined and the true standard. In the case of data like those just cited with the variable second, one might first conclude that the lower threshold was smaller than the upper, and then on the assumption that keenness of discrimination should be the same in both directions, draw the further conclusions that the standard or first interval must have been over-estimated. Such a conclusion would be erroneous. When the standard was given first it was as a matter of fact underestimated as shown by the fact that throughout the experiment the standard gave negative time-order errors when it was compared with itself. The above considerations indicate that the best procedure in studying time-order errors is to study separately the comparisons of each variable with the standard. In older studies of the time-order errors in the comparison of intervals, not infrequently the point of subjective equality was determined

[21] The temporal indifference interval determined by the method of mean error, *J. Exper. Psychol.*, 1934, 17, pp. 180–182, and Behavior with respect to short temporal stimulus-forms, *ibid.*, 1928, 11, Tables II, V, and VII.

by taking the average of the just noticeably longer and shorter intervals as determined by the method of minimal changes, without any consideration whatsoever of the vitiating effect upon this procedure of a skewed distribution. This faulty method of calculating the time-order error may well be one among the numerous causes of the notorious discrepancy in the conclusions reached concerning the temporal indifference interval.

Some interesting facts are also revealed by a study of the effect of practice on the limens. The various limens obtained with the group of 28 subjects have been approximately determined by calculating by simple linear interpolation the value of the variable giving 75 per cent correct judgments (including one half the equal judgments). The limens thus obtained in the case of the initial and final stages of practice, as represented by the data in Table II for the 1st 60 and last 60 trials, respectively, are shown in Table IV. The limens

TABLE IV

THE EFFECT OF PRACTICE UPON TEMPORAL DISCRIMINATION LIMENS

	Order	Lower Limen	Upper Limen	(U + L)/2
First 60 trials.....	V 1st	−12.7	+30.7	+21.7
	V 2d	−16.5	+ 9.4	+13.0
	V 1st and V 2d pooled	−15.2	+18.9	+17.1
Last 60 trials.....	V 1st	−15.8	+16.0	+15.9
	V 2d	− 8.6	+11.2	+ 9.9
	V 1st and V 2d pooled	−11.0	+13.4	+12.2

are given in the table as a percentage of the standard, though since the standard was 1000 σ, they may be read as absolute limens in units of hundredths of a second.

The data of Table IV show some very peculiar changes in the various limens with practice. Of the four limens with order kept separate (upper and lower limens with V 1st or V 2d) two grow larger (V 1st lower and V 2d upper) and two grow smaller. These changes are in perfect accord with the explanatory principles already advanced. According to these principles, the time-order errors are due always to a distortion of the first interval during the pause. Practice affects this distortion primarily by a gradual blotting out of the effect of

the absolute interval, and an ensuing predominance of the errors due to gravitation towards the series average. Now the effect of the absolute indifference interval under the present conditions is to cause an apparent shortening of the first interval during the pause. If, then, the first interval be the standard, practice will lessen the drop of this interval towards a shorter length, that is, increase the effective difference between the standard and a following shorter variable, and thus favor an apparent decrease in the size of the lower limen with the order V 2d. The effective difference between the standard and a following longer variable will at the same time decrease (thus accounting for the increase in the upper limen with the order V 2d). If, however, the first interval be the variable, the weakening of the influence of the absolute interval will cause the effective difference between a longer variable and the following standard to increase (thus in part accounting for the large decrease in the upper limen with the order V 1st) and will cause the effective difference between a shorter variable and the following standard to decrease (thus accounting for the increase in the lower limen with the order V 1st).

This tremendous effect upon the limens exerted by the lessening in the gravitation of the first interval towards a lower remote standard, presumably towards the absolute indifference interval, complicates the calculation of the effect of practice upon acuity as such. For it cannot be assumed that this gravitation affects all variables equally. An exact quantitative determination of the extent to which changes in the limens are due on the one hand to the time-order errors and on the other hand to a general improvement in acuity, constitutes an extremely difficult quantitative problem. Apparently, though, the four limens (upper and lower, V 1st and 2d) are more powerfully affected by changes in time-order errors than by change in acuity. An approximate indication of the change in acuity is afforded by the change in the average of the limens when the results with both orders are amalgamated (V 1st and V 2d pooled). The lower limen for the pooled data shows an improvement with practice from $-152\,\sigma$ to $-110\,\sigma$, or 28 per cent, and the upper limen a change from

+189 σ to +134 σ, or 29 per cent. The relative differential limen, averaging the upper and lower limens, improves from 17.1 per cent to 12.2 per cent. After the amount of practice here given, then, a difference of about 1/8th the length of the standard is necessary for correct discrimination in 75 per cent of the trials. This result is based on the pooled data for 28 subjects. The best of the subjects had a relative difference limen, calculated as above for the last 60 trials, of only 6.5 per cent, or about one fifteenth, while for the poorest subject the limen was 36.0 per cent, or over one-third the length of the standard. The limens for individual subjects, however, are very unreliable on account of the small number of trials from which they are calculated.

Summary

1. It is urged that the phenomena of temporal judgments are too important to be ignored in any theory either of time-order errors or of comparison in general.

2. Data on the comparison of empty intervals with a standard of 1000 σ reveal marked changes in the time-order errors during one forty-five minute sitting, and show that the changes occurring at one sitting carry over to some extent to a following sitting, so that the effect of successive sittings is cumulative. The shift in the errors is in the positive direction and is of sufficient magnitude to produce a change of several hundred sigma in the apparent indifference interval.

3. An explanation of this shift phenomenon is offered in terms of the same principles as used by the writer to explain the time-order errors observed in an experiment on weight-discrimination with a varying standard. According to this explanation, the second stimulus of a comparison pair is compared, not with the true magnitude of the first, but with a different magnitude called the effective standard. The immediate effect produced by the first stimulus is quickly modified by a tendency on the part of the subject to revert to a more chronic condition of attunement or expectancy. Calling these latter relatively stable conditions remote standards, the first stimulus may be said to gravitate towards one or more

such remote standards during the pause before the second stimulus. As a rule two such remote standards are influential in determining the value of the effective standard. One of these is an attunement to the average magnitude of the preceding stimuli set up by the experimental series and the other is some more permanent, though not unmodifiable, absolute standard. This latter standard in the present experiment is assumed to be the absolute temporal indifference interval, which presumably has a length in the neighborhood of 625 σ. In the case of other modes of comparison, such as sound and weight, it appears plausible that similar absolute standards may exist, though it may be that in some instances this standard corresponds, as claimed by Lauenstein, to the condition of no excitation at all.

4. The shift in the time-order errors with extension of the sitting or sittings is a result neither of a general increase in keenness of discrimination nor of a greater increase in keenness with the variable second than with the variable first.

5. The shift is readily explained by the assumption that as the subject becomes more perfectly habituated to the experimental series, the series-average tends to obliterate the influence of the absolute indifference interval; so that while initially the absolute indifference interval is the more influential of the remote standards, as the sitting is lengthened the series average becomes the more influential.

6. It is not a general rule of the successive comparison of magnitudes, as some previous investigators appear to assume, that with very short pauses (less than 3 seconds) between the compared magnitudes, the time-order errors are always positive. In the present instance, with a pause of only 2 seconds they were, initially, definitely negative. The explanatory principles already mentioned may be sufficient to account for the positive errors found in certain cases, as in Köhler's experiment on sound intensities.

7. The lower differential limen was found to be smaller than the upper differential limen. It is pointed out that this fact vitiates the procedure, sometimes used in calculating time-order errors (and from these, indifference values), of

subtracting the length of the standard from a subjectively equivalent value taken as lying halfway between the just noticeably greater and just noticeably smaller magnitudes.

8. As a result of the shift in the time-order errors with increase in the number of trials, practice sufficient to produce an average increase in acuity of about 29 per cent had the paradoxical effect of *increasing* both the lower limen when the variable came first and the upper limen when the variable came second.

9. The suggestion is advanced that the explanation of absolute temporal indifference intervals lies in a 'natural,' though extremely flexible and readily changed, adaptation or attunement for certain types of response on the part of the sensory-cerebral-motor apparatus rather than in the cumulative effect of the often repeated experience of some more or less constant interval of time, such as that required by the swing of the leg in walking.

10. The relative difference limen for empty intervals of time with a standard of 1000 σ, after moderate practice, was found to be, on the average, approximately one-eighth of the length of the standard, thus indicating a rather coarse human discriminal capacity.

[MS. received August 17, 1934]

THE CONCEPT OF RELIABILITY

BY RAYMOND ROYCE WILLOUGHBY

Clark University

Definitions of reliability usually turn out to be basically operational: reliability is (*a*) the correlation between comparable forms of a test, (*b*) the correlation between scores from the administration and re-administration of the same test, or (*c*) the correlation between comparable halves of a test, treated in such a way as to furnish a prediction of the corresponding value were either half to be expanded to the length of the whole. Interpretations, such as that reliability is the consistency of a test as a measuring instrument, the extent to which a test tells the same story twice, or the degree to which it is a representative sampling of the ability tested, are usually secondary. It is possible that this order, though often followed, tends to throw the emphasis in an inadvisable direction; it may be stimulating to consider the matter first from the point of view of the objectives which it is desired to attain, and second from that of the means suitable for attaining them. This is not to disparage the so-called 'operational' logic (a most useful point of view), which correctly understood maintains, not that concepts have no meaning beyond the operations invoked, but that they have no meaning in the absence of invokable operations.

Perhaps the best answer to the question as to what we seek in the concept of reliability is that we do not wish our 'true' results to be obscured by 'chance' factors. It seems to be implied that there exists in the subject a sort of noumenal trait or ability, in some fairly definite degree; and this is sometimes identified with the mean of an infinite number of measurements.[1]

[1] A procedure which appears to imply, somewhat paradoxically, that approximately half of any long series of measurements would indicate an ability better than the "true" ability. It might be better to identify the "true" with something like the

However, we may pass to the examination of the nature of the 'chance' factors affecting the 'true' ability. It may be observed at once that most of these are in the subject—or at least, none of them are in the test; a consideration which makes such a locution as 'consistency of a test as a measuring instrument' one to which it is very difficult to attach a common-sense meaning. The test, if the words are to be taken in their ordinary sense, can hardly be 'consistent'; it is sensibly the same test, just as a yardstick or a thermometer is sensibly the same yardstick or thermometer on different occasions, barring deforming accidents. However, it is easy to observe that there are tests which, applied to the same subjects at different times, do not give the same results; these so-called 'unreliable' tests, then, must be tests which contain some stimulus factor particularly likely to evoke different responses from the same subject at different times. Items ranging from unreliable to reliable in this sense are illustrated by the following:

Make a cross beside the name of the object you prefer:

vok	kumquat	arithmetic	ice cream
lum	mango	history	candy
gav	papaw	English	popcorn

In the first of these items, there is no ground for preference; the situation is such that preference cannot be tested in any clear sense, and the choice will depend upon vague and transient caprices of the moment—'chance' factors; in the last, preference is large relative to chance factors, and a change in marking it from one administration to the next presumably indicates a genuine change in preference. Obversion of the argument, however, must be undertaken with caution; it is not certain that a low degree of constancy from one administration to the next may be taken as evidence of lack of scope for the ability which it is desired to test—it may signify that the sub-

best ability demonstrable under conditions of strong homogeneous motivation, a procedure which should yield a minimum variation of specific performances, substantially all in the same direction, namely downward from the maximum. The matter is not quite comparable with physical measurements, for whereas one may estimate a linear measurement as longer than it is as a matter of chance, one could hardly estimate in similar fashion such an item, say, as the name of the capital of Turkestan.

jects have expressed a genuine change in themselves in this trait, that is to say, the trait itself is a changing one. This fact (which has been noted several times) constitutes the fatal objection to the use of the test-retest correlation as a measure of 'reliability'—it is composed of an inseparable mixture of (1) changes due to responses to two different stimuli (neither, possibly—as in the case of vok-lum-gav above—being that whose effect was sought) and (2) changes due to a genuinely changed response to the same stimulus. Thus 'telling the same story twice' is not necessarily the summum bonum in a test. It may be re-emphasized in passing that the elimination of (1) is the purpose of the determination of reliability (which, as just seen, cannot be accomplished by this means); the detection of (2) may sometimes be the object of study, in which case a different sort of test will be required from that necessary in cases where it is desired to examine some fairly stable trait (*cp.* disinclination for noisy games with reading ability); but in both cases the elimination of (1) is a desideratum. An interesting special case of (1) is the 'correlation between errors,' a bit of jargon signifying that, among other things, the memory trace of the previous stimulus, rather than the present one which it is desired to study, may be the effective stimulus.

The correlation between 'comparable' forms of a test is logically equivalent to the correlation between comparable halves. Both give some trouble. Kelley's (4) criterion for comparability (similarity without identity of elements) is frankly vague; there is no obvious way to tell at just what point elements become so similar that 'correlation between errors' results—that is to say, as before, a change or an identity in response cannot be separated with certainty into parts due to separate factors (in this case present evaluation and memory of former response). Without pretense at solution of this difficulty, however, it is suggested that the outstanding feature of the situation is *item intercorrelation;* random halves of a test, or forms which may be regarded as random samplings of a larger reservoir of items, correlate with each other because and to the extent that the individual items

are intercorrelated. The suggestion is made that all reliability of this sort reduces to this fact. It will therefore be appropriate to consider in some detail what is involved.

Concomitant variation between two items must, it seems to us, be presumptive evidence of allied causation.[2] It would be very difficult indeed to imagine a state of affairs in which two variables were genuinely correlated in the absence of all causal connection. The causal pathways, roughly, may be of three sorts: *A* may cause *B*, *B* may cause *A*, or a third factor *C* may cause both. Since in the usual test situation we hypothesize an 'ability' and endeavor to formulate items which will 'test' or manifest that ability, the most reasonable assumption of the three in the case of the intercorrelation of test items is that it arises because both responses are determined by the ability in question. It seems to us that such a proposition will warrant conversion: the presence of high intercorrelation between two items may be taken as evidence that both are representative manifestations of a common ability or trait—a unitary trait, if one wishes.[3] By induction,

[2] The problem of causation is a chronically vexatious one in psychology. In general our position is that set forth recently by Frank (2)—minus the implication that surrendering the concept of causation as a straightforward transfer of equal quanta of energy from member to member of a linear series necessarily makes the whole idea of cause an "episode" only. We think of the organism as a storehouse of potential energies the form and possibilities of which have been determined by the multitudinous influences exerted upon it throughout its past history. The fact that the releasing stimulus may transfer an amount of energy which is small with respect to the energy released should not obscure the consideration that it does transfer energy, and energy without which nothing happens; and this, as we understand it, is the best definition of a cause. Whether we wish to emphasize the releasing or some previous stimulus, or a combination thereof, is a matter of the requirements of the specific problem—it is only the concept of 'the' (*i.e.* sole) cause that is obsolete. We shall therefore speak of responses, *e.g.*, on personality questionnaires, as 'caused' by an underlying trait or tendency, meaning thereby that we think of the organism as having had impressed upon it by previous stimuli a tendency to react preferentially to a releasing stimulus. We are aware of the strictly behavioristic view that since nothing is observed but the reaction, it is illicit to postulate any intangible 'tendency' behind it; but this is a form of rigor that seems to place a premium upon sterile consistency as against common-sense usefulness.

[3] The problems of unitariness, scope, and specificity are closely interrelated. It seems to us that careful clinical observation, as well as common experience, clearly indicates the existence of uniformities in experience and behavior which may usefully be called traits. These entities are further described, but hardly disproved, by the

the presence of high average intercorrelation among items in a test—high reliability in the ordinary sense, revealed by correlating random fractions—may be taken as evidence that the items are manifesting a common trait. It should be noted that we have not disposed of the similarity-identity difficulty mentioned above, and that in consequence the argument is *apparently* vulnerable to a reductio ad absurdum, *viz.:* if we simply repeat the same item a number of times, we obtain perfect 'reliability' at the expense of some other desirable quality, perhaps 'validity.' Passing by 'validity' for the moment as having apparently no meaning beyond that of correlation with other criteria (often of equally doubtful trustworthiness) of the 'ability' hypothesized, we may observe that the undesirable factors introduced by repetition are

equally familiar facts that no gap separates them from the rest of behavior, that they appear in different settings, and that their uniformity is relative rather than absolute. In fact, the existence of adjectives and adverbs (implying that great numbers of individuals assign to and comprehend from them substantially the same meanings) seems to afford proof that human personality is not simply a congeries of isolated, specific reactions.

By unitariness we mean no more than that response items are associated with significant frequency, suggesting an organic connection. Since part of the argument of the present paper is that unitariness in this sense is a significant aspect of reliability operationally defined, a critic has suggested that 'reliability' of tests (*e.g.* 'omnibus' tests) known to be non-unitary constitutes a case requiring special attention. The solution appears to be that such tests, *as tests*, are not reliable; their high reliability coefficients are obtained by pairing items between forms, and so are not a random sampling of correlations between all possible halves; thus the test is rather a group of probably reliable sub-tests than a reliable test in its own right.

The scope or breadth of a trait we find a difficult concept. Students who speak of it appear to have in mind a continuum, in which different traits have breadths of all possible degrees; such a concept seems to follow from a very empirical use of the term *trait* as, perhaps, a general term for any attribute capable of being named (including, *e.g.*, fear, fear of animals, fear of spiders). It is then maintained that samplings of equal size will be representative of the traits in proportion to their narrowness, so that the test will be more reliable for the narrower traits. We favor a somewhat more naturalistic concept of trait (possibly one similar to 'general factor') in which the boundaries lie approximately where the manifestations cease to be significantly associated with the core or with the constellation as a whole. In the former case the increased reliability of the narrow "trait" would reflect rather the fact that fewer genuinely different manifestations of it were possible than in the case of the broader one (which would conform to the conception herein advocated), so that instead of correlating manifestations we should really be correlating identities. Thus there are fewer genuinely distinct manifestations of fear of spiders possible than of fear of animals, because there are fewer spiders!

primarily of the psychological and only secondarily of the statistical order; if we could be certain that the subject would give his careful attention to the answering of the item each time it was repeated (if each time, say, he were to be afflicted with amnesia for the previous times) we might indeed consider the evidential weight of his response to have been increased by the repetition, just as we suppose a case to be strengthened by repetition of the same testimony from additional equally reliable witnesses. However, the subject does not forget that he has just been asked the same question, and the effect is to introduce factors of resentment and boredom, which are chance factors in the sense discussed above and tend to decrease, not increase, the reliability.[4] The relationship of this aspect of the problem to that of scope discussed in a footnote above will also be evident.

This conception of split-half or comparable-form reliability as simply inter-item correlation can and should be brought into relationship with Kelley's (3) concept of reliability as adequacy of sampling. Here we have, as before, an hypothesized noumenal 'ability' lying behind all manifestations of it, and the desideratum is that these manifestations shall be 'representative.' This term usually supposes a continuum of some sort from which samples are taken of equal size and at equal intervals; thus a representative sample of a population in stature would be formed by arranging the entire population in order of stature and selecting every tenth individual. In what way, however, an ability presumed to exist in a single individual can be equated with such a continuum is not clear, especially if we accept the suggestion advanced above that the 'true' ability be identified (approximately) with the best ability. Additional light on the matter may be obtained from the corollary that the well-known increase in reliability due to increasing the number of items (the fact at the basis of the 'prophecy formula') is presumed to imply that the additional items make the sample more representative. This is still difficult to understand unless in some way we can

[4] Further, in any practical testing program we do not wish to take up his time, or our own, with mere duplicate questions.

think of an individual's ability as a collection of elements spatially distributed; but if we ask ourselves in what way the observed effect *could* understandably be produced, we observe that since the added items (as well as those already present) were included in the belief that they reflect the ability being tested, they will ordinarily correlate with that ability; thus the number of items contributing to the average item intercorrelation is increased and its coefficient correspondingly raised.

This brings us to the possibility of making much simpler and more direct the task of obtaining good test reliability; since the object of having such reliability is the elimination of responses to irrelevant stimuli, and since intra-correlation is evidence that such responses are few, it will be preferable, rather than merely adding items in uncritical obedience to the Spearman-Brown formula (in the hope that some will increase r_{11}), to add only items which empirically show good correlation with the total of those already present (and, incidentally, to eliminate any already present which conspicuously fail to meet this criterion). Although we have heard expressions of pity for investigators who 'thought they could make short reliable tests despite the explicit statement of the Brown-Spearman formula that it can't be done,' our native reverence is unfortunately insufficient to sustain us in the face of the fact (*e.g.*) that Thurstone's well-known neuroticism inventory has been reduced by Bernreuter (1) to approximately one-half, and by Willoughby (7) to one-ninth, its original length without material loss of reliability.

The similarity-identity impasse has, however, been urged against these attempts. It is said, for instance, that there is something artifactual about them—that the process of selecting items highly intercorrelated overloads the resulting scale with duplicate or near-duplicate material. There is a certain circularity about this argument—it is first alleged that items correlate only by reason of duplicate elements, and then concluded that since these items correlate they must be duplicates. We doubt our ability to break ground here much in advance of that traversed by Kelley in his criteria for comparable tests; that unanalyzed complex of bases of judgment loosely desig-

nated as 'clinical' must still, it appears, comprise the major ground for selection. Nevertheless we are inclined to remain in some doubt about the concept of duplicate or identical elements; the problem of generality *vs.* specificity is again involved here, and we know of no very adequate or fundamental analysis. Empathetic observation of human beings systematically and protractedly trying to give expression to internal events has a way of engendering, at least in the personality sphere, certain generalizations of an *als ob* type, *e.g.:* it is as if these people were activated by a generalized drive which could be called a fear of a certain class of objects or situations, or one which could be described as an impulse to attain more intimate association with objects of a given class. It appears to us that these are perhaps primitive versions of genuine scientific hypotheses; but we do not feel ourselves equipped to contest the point with anyone who wishes to maintain that inasmuch as they include neither centimeters, grams, seconds, nor standard deviations they are not science but mysticism. In any event, possibly the drives they postulate may be regarded as identical elements inherent in the different observable manifestations; thus a fear of snakes and a fear of spiders may have as an identical element some more or less generalized fear, and to this extent we may perhaps grant that the items are duplicate or identical. However, it appears to be this fear, or something supraordinate to it, such as neuroticism, that we wish to measure, and it seems allowable to do so by testing its presence in different manifestations, on the supposition that its general presence may be better inferred from exploratory samplings (in which sense this term may perhaps be brought into relationship with Kelley's use of it) at several points of possible emergence than at a single one. On this hypothesis we may postulate, as the best criterion of which we are capable at the moment, that items may be regarded as different when they sample different manifestations of the same trait (as fear of snakes and fear of spiders) and duplicate when they sample the same manifestation (as fear of spiders and timidity about arachnids); in other words, to

qualify as separate items the difference between two elements must be situational, not merely verbal.[5]

A few words may be essayed about the formula so widely used for estimating the reliability of a total instrument from the intercorrelation of parts of it. From our present point of view, while the formula is impeccable in its derivation from its premises, these latter are rather too abstract to represent with maximum usefulness the usual practical situation in which they would naturally be invoked. Since, as in most statistical derivations, argument must take place in terms of additive identical units (or not at all), the important datum must perforce be neglected that real test units are by no means all identical.[6] Thus the very fact that an increase in reliability will occur when a test is lengthened depends upon the complete similarity of the added material, at least on the average, to that already present; and the degree to which the formula will actually predict the amount of this increase depends upon the conformity of the material to this assumption, a fact which appears to be fundamental in the problem, now under investigation, of the applicability of the prophecy formula to learning data. But, as indicated above, only the virtue of docility requires observance of this precaution; if it be really desired to increase the reliability of a test, those items may readily be selected which will increase it (and without any lengthening whatever) even though it be known that they are incomparable with the items already present. If, however, the items have already been made highly homogeneous with respect to the amount they contribute to the total, so that they are maximally and psychologically as well as statistically

[5] Note, however, that work on the form of questions in testimony and in questionnaires (v. Smith, 5) suggests that even a verbal restatement may constitute a somewhat different situation.

[6] Hence, probably, the uncomfortable feeling of many clinical observers in the presence of statistical evaluation that the baby is being thrown out and the bath carefully studied; babies are in fact notoriously refractory to complete description in terms of equal units. Specifically, it must be admitted that many tests suggest strongly that the necessity to arrive at statistically acceptable (meaning principally reliable) results had gradually forced a change of objective in the course of the work, namely from measuring a significant trait that proved statistically intractable to measuring one (often identically named) from which most of the significance had been lopped away to obtain equality of units.

equal, then the formula may yield an expression for estimated 'reliability' of the whole test that is more than a numerical statement.

The relationship between reliability and validity has not been extensively discussed in print, possibly for no reason more abstruse than that it is generally understood. The problem, if any, appears to reside in the consideration that correlation with a criterion may be regarded as a special case of correlation between test items, and that the two functions of reliability and validity, instead of having the discrete character implied by the undergraduate definitions, are really continuous.[7] Thus on the more obvious level it is said that a psychotechnical test has a split-half reliability of .7 and a validity, as determined by correlation with foremen's judgments, of .6; and it is estimated that but for the 'unreliability' of both test and foremen's judgments this validity, or correlation between 'true score' and 'true criterion,' would have been .9. In such a derivation it seems to have been assumed, as before, that the foremen's judgments may be thought of as some sort of continuum of equal and additive elements, the addition of more of which will result in more representative sampling of the continuum and hence increased 'reliability.' We submit, however, that such judgments are in essence only another test item—one drawn, to be sure, from a region of quite different forms of manifestation, but still an effect of the same presumptive ability which causes the subject's test responses; the ability, so to speak, has manifested itself once in

[7] This is Thurstone's (6) point that validity is the consistency of a new measure with old ones, while reliability is the consistency of any measure with similar ones; it seems also to be the kernel of considerations grouped around the word *prediction*. Any test, it is said, is a sample of behavior from which it is hoped to predict further behavior, and the minimum essential for this is that the sub-samplings of the sample predict each other with fair consistency (reliability), as well as that the sample as a whole predict other, preferably slightly different, samples (validity). While no outright objection can be made to this view, it has a flavor of unnecessary rigor—as if to assert that no hypothesis of the dynamics of the behavior is desirable. From our standpoint, the assumption of, *e.g.*, differential susceptibilities to stimulation by different classes of objects has some probability, and its implications are of even more significance than the raw fact that response *A* is accompanied by response *B* in *r* per cent of the cases. We feel, in short, that procedures looking toward an organismic description are preferable to ones limited to accounts of particularistic behavior.

each test response and once in the foreman's consciousness and overt judgment. It is well known that such judgments are not remarkable either for their reliability or for their correlation with test results; and this we should expect from the discussion above—an item which consists of the reaction of another individual to a particular phase of a subject's activity is particularly susceptible to influence by irrelevant stimuli. Thus a foreman, requested to evaluate the mechanical ability of John Smith, may on one occasion quite unwittingly react in terms of the latter's recent insubordination and on another in terms of his habits of neatness; either response will decrease the validity of the mechanical ability test under investigation, and both together will decrease the reliability of the criterion.

It is suggested that the foregoing remarks on item intercorrelation emphasize a fairly promising mode of approach to the problem of the 'unitariness' and genuineness of traits, which, as we understand it, is the intention of the concept of validity.[8] A homogeneous, closely knit test, in which each item is highly correlated with the others and thus makes a strong contribution to the total, is to be suspected of measuring, with them, a really existent trait; although it may not measure the trait we wished, it is nevertheless apparent that we have succeeded in selecting from all possible human responses a group which forms a natural constellation. Such a trait would be reliable in the sense of having high inter-item correlation, and valid in the sense of genuine or corresponding to some psychological reality. Of all remaining responses, there are no doubt some which, if we knew them, would be found to correlate highly with these and thus to belong in the trait; others would fail to correlate, but might belong in other constellations (*e.g.* intelligence vs. neuroticism, which appear to be two independent constellations); still others would correlate more or less slightly, and the presumption in such cases would be that they yielded 'unreliable' responses, re-

[8] In making such a suggestion it is not intended to minimize the factor-theory contribution; its techniques are incomparably neater and more powerful than the pedestrian ones suggested here, though its logical and psychological connotations still baffle us.

sulting perhaps on several occasions from the relevant stimulus, but on others from irrelevant ones.

It remains to summarize the several arguments presented. The chief of these is that the principal object of reliability determinations is the detection and elimination of items especially susceptible to response determination by stimuli other than those under study. Such an item may frequently be detected and improved by empathetic clinical analysis or by actual introspections of subjects; where these fail or are impracticable, statistical indicators may be sought. Among the latter, marked or irregular temporal variation of response may be useful, but is subject to the important objection that it is to an unknown degree the product of genuine changes in the stimulating power (*i.e.* the amount) of the presumptive ability or drive being measured. Most important, however, is the degree of correlation of the item responses with this ability or drive; this cannot be observed directly, but its strength may be inferred from the consideration that intercorrelated items must have a common determinant and must be more highly correlated with this determinant than with each other. A highly reliable test, therefore, is a test the elements of which are highly intercorrelated; and it turns out also to be a highly valid test in the sense that it is 'unitary' and therefore genuine. This argument, that susceptibility to response determination by irrelevant stimuli results in lowered intracorrelation, appears to be contraposable, *viz.:* high intracorrelation implies response determination by relevant stimuli predominantly. Accordingly, the preferred technique for effecting increased reliability is not blindly lengthening the test but selecting items maximally intercorrelated. The possibility that an item may be highly intercorrelated with others merely because it is a duplicate may be settled practically by examination of its content from a psychological point of view: if its difference from others is situational, *i.e.* if it represents a different manifestation of the same ability or drive, it is to be regarded as a distinct and legitimate item; if its difference is merely verbal, *i.e.* if it is an identical manifesta-

tion of the same ability or drive, it is to be regarded as substantially a duplicate.[9]

REFERENCES

1. BERNREUTER, R. G., The theory and construction of the personality inventory, *J. Soc. Psychol.*, 1933, 4, 387–405.
2. FRANK, L. K., Causation: an episode in the history of thought, *J. Phil.*, 1934, 31, 421–427.
3. KELLEY, T. L., The reliability of test scores, *J. Educ. Res.*, 1921, 3, 370–379.
4. ——, Statistical method, New York: Macmillan, 1924.
5. SMITH, R. B., The measurement of inferiority feelings at the high school level, *Psychol. Bull.*, 1932, 29, 633–634.
6. THURSTONE, L. L., The reliability and validity of tests, Ann Arbor: Edwards, 1932.
7. WILLOUGHBY, R. R., Some properties of the Thurstone personality schedule and a suggested revision, *J. Soc. Psychol.*, 1932, 3, 401–424.

[9] Acknowledgment is gratefully made for helpful criticism to Drs. A. Anastasi, M. N. Crook, C. H. Graham, W. S. Hunter, D. E. Johannsen, V. A. Jones, and F. Linder.

[MS. received August 20, 1934]

PERCEPTION AND IMAGINATION

BY KATE GORDON

University of California at Los Angeles

Persons on the witness stand are often reminded that they must tell just what they saw and not what they thought. It is assumed that the witness knows the difference between his perceptions and his conclusions. We all admit that there is such a difference and that it is an important one; but it is far from certain how the line should be drawn between what is actually present to sense and what is a figment of mind. This criterion—sensuous presence—is used in psychological literature to define the distinction between perception and imagination. These two are defined as if correlates. Thus perception is said to be an awareness of objects present to sense. Imagination is said to be awareness of objects not present to sense.

A consideration of this criterion raises the question: what is it that is present to sense, *i.e.* what classes of objects are perceptible. Also the question whether there are other objects which are imaginable, and how the imaginables and the perceptibles are related to each other.

The relation between perception and sensation requires some comment. Perception is sensation apprehended or understood. A convenient phrase gives it as 'sensation plus meaning.' In each of these descriptions a close alliance is indicated between perceiving and sensing. But there are certain differences. (1) The sensations are occasioned by the stimulation of special sense-organs. Their qualities are always correlated with specific nerve processes, and we cannot get the same qualities through different sense organs. But in perception we frequently identify a single object through several different organs. (2) Another difference is that the sensations, such as yellow, sweet, middle C, warm, etc., have

an abstract and 'adjectival' character, that is to say, they appear to us as attributes of something else, and not as independent objects. In case of perception, on the contrary, we apprehend objects, and these are often denoted by nouns.

Perhaps it is these differences which have suggested the idea that in sensation we are given something which is subjective, whereas in perception we are given something objective. But it is not necessary to assume this difference. It seems better to say that the sensation is merely the simplest case of perception. A sensory quality of yellow which we get when looking at a yellow sunflower is more abstract, more adjectival than the sunflower; but it need not be more subjective. Sensations *are* something and they are *of* something. We do not take them to be spontaneously generated by our minds. Sensations are the perceptions of simple qualities.

In every percept, if the preceding comments are correct, we recognize two aspects or elements. These are the sensory and the significant. We can take some common object for illustration, say a dish of soup served for a family. As the odor of the soup floats through the house it acts as a signal to the different members of the family group. The dog wags his tail, the children ask how soon dinner will be ready, father looks at his watch and mother powders her nose. Each of these reactions is a mode of interpretative behavior. Each reaction shows that the sensory stimulus was understood. When the food is served and eaten there are many other sensations from it, vision, touch, taste, temperature, muscular sensation, perhaps sound. But all of these sensations are referred to a single object, or thing signified. Each person who partakes of the soup refers his individual sensations to it, and there is a group reference or understanding by which the members of the family discuss the food as if it were a single and a common object.

A quotation from Thomas Davidson,[1] gives a vivid statement of the contrast between the intellectual element and the sense element in such a perception. He says:

> Intelligence—of which religion and science are the two chief

[1] T. Davidson, Education as world building, *Educ. Rev.*, 1900, 20, pp. 225–245.

manifestations—as distinguished from sense, consists just in placing something behind, or under, our sensations. This act is called hypothesis, hypostasis, or supposition; that which results from the act, essence, substance, subject, idea, reality, cause, thing-in-itself. . . . Strictly speaking, these essences or causes are not known, in the sense in which sensations are known. They are objective, whereas sensations are subjective.

Aside from the idea that sensations are subjective and perceptions objective, to which we took exception above, this is an admirable statement. Every percept involves hypostasis. Even simple sense qualities are things. The answer to our question, what are the perceptibles, may begin therefore with these.

(1) Sense qualities themselves yield a series of categories of perceptibles. We perceive color, light, sound, taste, etc., etc. All such discriminable differences are perceptibles.

(2) Next may be named certain attributes of sense experience which give us perceptible differences. The first of these is intensity. It occasions such judgments as 'sweeter,' 'louder,' 'darker,' etc. These differ from quality judgments in that they report degrees on a scale.

(3) Another attribute, duration, affords us another range of perceptions. Whether we can perceive unfilled time may be disputed. But there seems no doubt that (within certain ranges) we have immediate awareness of longer and shorter intervals of time.

(4) Spatial extent is a third attribute commonly allowed to afford us perceptions. Besides these four attributes of sense, namely, quality, intensity, duration and extensity, there are other aspects of perceptual experience which may be less fundamental, but are none the less genuine. These are (5) number, (6) shape, and (7) movement.

Now in addition to the kinds of things mentioned above we come to classes of objects which give rise to sharp differences of opinion as to whether they are perceptible or not. Can we, properly be said to perceive another person's feelings? Is it correct to say that we can see his anger or his fright?

Concerning the possibility of perceiving the emotion of

another person, Professor Helson defends the position that it is possible to see the expression of a face without being first explicitly aware of the details of color, light, etc. which make up the sensory qualities of the image. He writes:[2] "Can one believe that my perception of what we may call a 'friendly face' is adequately described as being merely a concatenation of so many hues, tints, and chromas, having such and such an extent, intensity, clearness and lasting for so many seconds?"

In answer to this query we may agree that there is more in the perception of a friendly face than the sum of its sensory elements. Probably there are memories and expectations. But we might contend that what we actually see can be only the qualities which are fitted to stimulate the eye. If we happen to be a painter trying to represent a friendly face on canvas, we do have to reduce it all to just those items of hue, tint, chroma, intensity, extent, shape. If the expression of a face is to be *conveyed*, it has to be conveyed in the terms of a particular vehicle or medium. Nevertheless, the extent to which expressions of emotion by face and posture can be conveyed to an observer must be decided by experiment. We have the work of Landis, Feleky, Langfeld and others to throw light on this. It has in fact become a routine class experiment to exhibit series of photographs and to call for opinions as to what they express. Such results show that there is agreement on some pictures and disagreement on others. These facial expressions very probably have a native basis, like language itself, but they become modified by convention. The reading of feeling in a face then becomes the understanding of a set of somewhat arbitrary symbols. It is easy to learn some of these symbols. It is possible also to learn to read signs of feeling in the face of an individual whose reactions may be different from the average, but consistent with himself. I think we must conclude that there is no infallible or necessary connection between the facial pattern and the emotion which it signifies. But we must also recognize that such facial expressions, or gestures, are often correct indices. What they give us is worth some risk.

[2] H. Helson in Wheeler's Readings in psychology, Crowell, New York, 1930, p. 394.

Another question of a like nature is, can we perceive a person's character traits by inspection of his face and form? The results of tests where photographs are used seem even less promising than the results on judgments of emotional expression. The pictured faces are not safe guides. We cannot, however, say that opinions based on photographs are entirely worthless. But their accuracy is nothing like that of the perception of the physical features themselves. For example, we cannot very well perceive by a person's eyes whether he is honest, but we can quite well perceive whether his eyes are light or dark, large or small, near together or far apart, and such like physical facts.

A third question is, can we perceive a human act? We do speak of observing human behavior, as if it were a scientific method. And the witness in court is allowed to say that he saw a person smile, though he must not say that he looked happy. Granted that we may perceive the movements of a person, does this mean that we perceive also the significance of the movements? The following illustration shows our difficulty.

A teacher of mathematics accused two young men of cheating in the examination. She said to them: "I saw you do it." And she described how one had looked over on the paper of the other. The boys denied the charge, and the one whose eyes had been turned towards his neighbor's paper declared that he had seen nothing. The instructor was unconvinced until finally she read the two papers and found them to be entirely different. There is more than one lesson to be drawn from this episode, but for our purpose it is enough to say that certain movements had been perceived, but no act of cheating. Soon afterwards, this writer was giving an intelligence test to a large group of students. I saw what appeared to be a case of cheating, but I was determined not to make the mistake of the mathematics instructor. I watched the suspected one, and saw her glance repeatedly at her neighbor's paper. Finally she leaned over, planted her finger on the paper and began talking. This seemed too flagrant to overlook. I went down and asked what the trouble was, and the student explained

that she was only a visitor and that she had spoken to her friend only after the latter had finished her test. This put an entirely different construction on the series of movements which I had observed. In each of these cases the person's motions were seen but in neither of them was the moral quality rightly apprehended.

What we have said expresses doubt as to any necessary connection between sensuous patterns and such things as human feeling, human character, and human conduct. A classical passage relevant to this point is found in Hume [3] where he says:

> There is no object which implies the existence of any other, if we consider these objects in themselves. . . . It is therefore by *experience* only that we can infer the existence of one object from that of another.

If we apply this principle to our question it will lead us to deny that there is any intrinsic association, *i.e.* any necessary connection, between the sense patterns and the objects which we take them to signify.

We have, however, an institution which has for its major purpose the exhibition to sense perception of human character, emotion and action. This institution is the theatre. Here, if ever, we find human quality made patent to the senses. And not drama alone, but the arts, generally, strive to give to the observer sensuous images of life. Hamlet says of the play that it is the image of a murder done in Vienna. In the phrase of Hegel, art is the manifestation of idea to sense.

But how does art operate to make manifest to our senses those features of life which perhaps are, in their full reality, supersensible? It seems to do this by the selection of sensuous items which are peculiarly appropriate as vehicles for the thought of the artist. This statement assumes that there are in fact some items which are 'peculiarly appropriate' or more appropriate than others to the rendering of the ideas in question.

Writing of such matters George Moore [4] has said:

[3] D. Hume, A treatise of human nature, pt. 3, sec. 6, p. 86 of Everyman's ed. 1911 (first ed. London 1738).

[4] G. Moore, Confessions of a young man, New York, The Modern Library ed., p. 148 (first ed. 1888).

> The symbol, or the thing itself, that is the great artistic question. In earlier ages it was the symbol; a name, a plume, sufficed to evoke the idea; now we evoke nothing, for we give everything: the imagination of the spectator is no longer called into play. In Shakespeare's days to create wealth in a theatre it was only necessary to write upon a board, 'A magnificent apartment in a palace.' This was no doubt primitive and not a little barbarous, but it was better by far than by dint of anxious archaeology to construct the Doge's palace upon the stage. By one rich pillar, by some projecting balustrade taken in conjunction with a moored gondola, we should strive to evoke the soul of the city of Veronese: by the magical and unequalled selection of a subtle and unexpected feature of a thought or aspect of a landscape, and not by the up-piling of extraneous detail, are all great poetic effects achieved.

The artist, according to this fine passage, creates, by his 'magical and unequalled selection' new perceptibles for us.

One more class of objects is to be mentioned, these the most doubtful of all as respects their perceptibility. Do we ever perceive abstract ideas? Can we ever see 'horse,' 'man,' 'animal'? Can we ever come face to face with justice, truth, beauty? Or have these things, indeed, no faces, no perceptible aspects? The following observations seem relevant to this question.

Reading experience shows us that we may become conscious of certain wholes without being at the same time conscious of their parts. We see words rather than letters, and even phrases rather than words. At least this is so of the practised adult. It requires practice to see the letters, as we know from proof reading. We can read words and phrases at a greater distance than we can read the letters. A further fact about our reading is that we can get the meaning from a printed page without being conscious of the precise words which carry that meaning.

Experiments in sense discrimination show that we can perceive the fact of difference in cases where we cannot perceive the direction of that difference. Listening to two tones, we can at times tell that the two are different in pitch, but without knowing which is the higher. When a contact point on our skin is moved, we can report accurately the fact of

movement before we can tell whether the movement is up or down, left or right.

As a person comes towards us from a distant place we first see something in motion. This something develops into human shape. Finally an individual is recognized. This illustration is used by Hamilton, in the 'Lectures on Metaphysics,' and he concludes the passage in these words:

> Thus . . . children, first of all, take note of the generic differences, and they can distinguish species long before they are able to discriminate individuals. In all this, however, I must again remark, that our knowledge does not properly commence with the general, but with the vague and confused. Out of this the general and the individual are both equally evolved.

In each of the cases cited above we assume that the differences mentioned must *be* particular differences, even though we do not sense them as such. A word cannot be present on a page without being made up of individual letters. A person must *be* a particular person even though we do not recognize him. A tone, in order to vary from another in pitch has to be either higher or lower. A movement must be in a determinate direction. The situation is that we suppose ourselves in the presence of particulars but we perceive only the general aspects of them. Lossky (The intuitive basis of knowledge) believes that we rarely attain to a knowledge of anything in its full particularity.

Berkeley professed that he could not frame an abstract idea of a triangle, unless it were a particular kind of triangle. Locke thought that he could think of triangle in general. Hamilton says of Locke that he held this view in 'all its revolting absurdity.' Now it is true that you cannot *make* a triangle unless you make it isosceles, or equilateral or scalene. But you may talk about triangles, and make propositions about them without being limited to a single specimen. And I believe that we can perceive that a figure is triangular without being sure whether the sides are exactly equal or not. Let someone hand you a triangular card while you keep your eyes shut. You can assure yourself by running your finger

over it that there are three corners. But can you be sure of the exact length of the sides?

Discussions of mental imagery often mention the so-called generic image. The composite photograph is a convenient analogy. In the photograph, the particular images, falling on the plate, result in a blurred image in which individual features are cancelled and common features emphasized. There are two ways of treating such an image. One is to regard it as nothing but a particular image after all, inasmuch as it has color and contour and a definite location in space and time. The other way is to be mindful of how it was formed and therefore to see it as a better symbol of the abstract idea. It is something like a master key, which is of course still a particular key, but which gives a more comprehensive performance than other particular keys.

The status of general ideas has been described in the great historical controversies in a variety of ways. Sometimes they have been viewed as antecedent to the particular things of which they are predicated (realism). Sometimes they have been viewed as derivatives of the particulars (nominalism). Again they have been held as inherent in the particular (conceptualism). Now it seems as if each of these views must be true part of the time.

In the case of manufactured objects we have no trouble in supposing that the general idea of a thing precedes the particular embodiment of it. Somebody thought of automobiles before there were any. The process of manufacture is an embodying of the idea in particular forms. We cannot make things in general. All making is making specific.

The position of nominalism, that the abstract ideas are made by us after we have come in contact with the particular facts, is a fair description of what does indeed happen at times. For example, a mass of data gathered by a laboratory assistant may be handed over to another person for interpretation. Conceptualism also offers this valid notion, that the same thing may be regarded as a symbol or specimen of a class, and also be regarded as an individual. I cannot help believing that each of these views truly describes certain moments of ex-

perience, the universal before the thing, the universal in the thing, the universal after the thing.

Locke[5] says concerning abstract ideas: "They are the Workmanship of the Understanding, but have their foundation in the Similitude of Things." Does, then, the relationship of similarity exist in the same way that the objects exist which are similar? For my part I do not doubt it. It is true that if you were to dissect a dog you could perceive his bones and muscles and lungs and stomach, but would not find anything which you could name his similarity. Yet his similarity is a fact, and if you were to dissect two dogs, then you could perceive that similarity in his bones, his muscles, his lungs and his stomach. It is 'there' just as much as the two dogs are there.

Ideas are constructions and they are also discoveries. As an example of a general idea which is both a construction and a discovery the following is offered. Figure 1 shows the

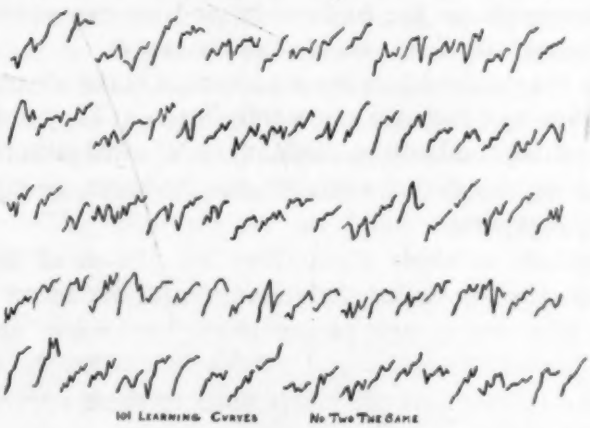

FIG. 1.

individual learning curves of 101 persons. An experiment was performed in which they were asked to write, upon signal, in ten seconds time, as many letters of the alphabet as they could, in the reverse of the usual order. Ten such trials were given and the curves plotted. No two of these curves are exactly alike. But if we calculate the median number for

[5] J. Locke, An essay concerning human understanding, Bk. III, ch. 3, sec. 13, London, 1690.

each trial and construct a curve from these numbers, *viz.* 7, 9, 11, 13, 14, 14, 14, 15, 15, 16, we shall have a curve which represents the performance of the class as a whole. This curve will not coincide with any single one of the curves given in the chart. It is not a literally true record of anybody's learning. It is an artifact, and there is one respect in which it gives a really false impression. It fails to show any retrogression, whereas in 91 out of the 101 individual curves there is some trial which shows a worse score than the one which preceded it. In other words, the results show that it is normal for members of the group to have irregularities in progress. But inasmuch as the retrogression appears in different parts of the curve for different persons, this appearance is cancelled in the general curve. Nevertheless the general curve is not a baseless nor an arbitrary fiction. It is a discovery; for it records the facts about the progress of a class which were not discerned without this curve. We can make true statements on the basis of it, and we can predict what other classes will do in similar circumstances.

Such graphic symbols are attempts to make abstract ideas perceptible, and they are not wholly futile. They are sensory shapes which embody meanings in a semi-pictorial way. They do require verbal explanations, however, to make their meaning complete.

Linguistic symbols themselves are the most important means that we have for understanding or for using abstract ideas. The words are perceptibles, but what about the objects for which they stand? Must we not acknowledge a limit here to the class of objects open to sense apprehension? Plato in the Statesman says:

> But people seem to forget that some things have sensible images, which may be easily shown . . . while the greatest and noblest truths have no outward image of themselves visible to man, which he who wishes to satisfy the longing soul of the inquirer can adapt to the eye of sense, and therefore we ought to practice ourselves in the idea of them; for immaterial things, which are the highest and greatest, are shown only in thought and idea, and in no other way, and all that we are saying is said for the sake of them; moreover, there is always

less difficulty in fixing the mind on small matters than on great (Jowett).

We have to acknowledge with Plato that these highest and greatest truths are not wholly on the perceptual level. But again we feel that any truth which is at all communicable has to make use of perceptible means of communication. And any man who deals justly must finally exemplify justice in perceptible acts. Physical instruments, such as the telescope, microscope, telephone and radio, have enlarged the field of perceptibles beyond all expectation. It would be dogmatic to assume that, *e.g.* justice must always remain an imperceptible. I believe it is Holdworth who has suggested that the traditional forms of the law are so many efforts to dramatize justice. It has already been said that artists strive to give the essence of things in sensuous form. There are times when we feel that they succeed and that there is nothing greater and higher than what they give. Finally we may ask how it is possible for anything spiritual to be shown in material form, how, that is, shapes or sounds or words or dramas or ceremonials can hold and transmit that which is moral. We may not know how this is done, but we can be sure that it is done. The inner world of moral values must somehow and at some time become geared with the external world or lose its integrity. The abstractions all have connection with the perceptible world.

In the preceding discussion of perceptible objects we have assumed that these objects have existence which is to a large extent independent of the particular individual who perceives them. But at the same time we assume that every perception is the act of an individual person. The object is always revealed through individual views or perspectives. And the same object may be revealed through perspectives which are quite different from one another. If we want to understand what happens in perception we have to study, not only the object perceived, but the perspective through which it is grasped. In an earlier paragraph the work of the artist was mentioned as a magical selection of items by which his ideas were to be shaped for our senses. Now we must note that in

the act of perception itself a similar work is performed. The percipient sifts and shapes, whether consciously or not, the parts of the object which he observes. These selections, distortions, or personal views are the individual perspectives. This is why we class perceptual images among the products of imagination.

The word 'perspective' is most familiar to us in the field of the visual arts. Certain features of such 'views,' 'projections,' or 'perspectives' are matters of common remark. Parallel lines seem to converge in the distance. Near objects look larger than distant ones of the same size. The conventional signs of depth and distance belong to these individual views. A group of persons drawing likenesses of the same model will produce sketches which are different, for the reason that no two persons can occupy the same position at the same time. Persons seated in a theatre see the stage from points of view which are different. No two seats, hence no two views are identical. There is nothing peculiarly psychical about such perspectives. They are recorded by the camera as well as by the eye. Nevertheless, at the moment when they occur, they are unique perspectives or individual views. We may name them perspectives due to space location.

A second class of perspectives is due to the character of the receptors, that is, the sense organs by which impressions are received. The person who is color blind or color weak perceives in peculiar ways. The world of the totally blind or the totally deaf is, in its range of qualities and in its emphases, profoundly altered from the general. An interesting class of cases are those in which one sense mode represses or distorts the testimony of another. An example is the size-weight illusion. Thus when we lift two objects which really weigh the same but differ in size, we feel the smaller to be the heavier. Through vision we are 'set' for a lighter object and this misleads us in our perception. We had a misleading perspective.

A third class of perspectives is due to the principle of contrast. This is probably not inferior in importance to the principles of space location and of sense organ structure as a source of perceptual differences. It is clear that without some

contrast between an object and its surroundings there can be no perception whatever; for in order to be discriminated an object must be discriminated from something else. We get an object in different perspectives when we view it against different backgrounds.

A fourth principle, which is a determiner of perspectives, is habit. This operates in two seemingly opposite ways to affect our perceptions. On the one hand there are cases in which attentive practice will increase the reliability of our perceptions. Proof reading is an example. But there are other occasions when an habitual association may overrule us and make a percept inaccurate. One time at an afternoon tea I was looking into a cup and noticing the rich amber color of Orange Pekoe tea. I took a drink of the beverage, and, to my surprise it proved to be a claret punch. Immediately the color seemed to change and to become claret color. The apparent change in the color was very astonishing and disconcerting. The same experience was mentioned by another guest. The expectation of getting tea had been created by the wording of the invitation and also by the shape of the container in which the liquid was served. A large number of such reports have been published by Grace Adams [6] and she has arranged ingenious experiments in which just such changes in the apparent sensory qualities have been brought about by means of ideational suggestion and the appeal to habitual modes of perception.

Other illustrations of the perspective of habit are found in the tendency to see known objects as of some constant size. When we hold our own hand close to the eye and then slowly move the hand out to arm's length, we do not see it shrink in size to the same extent that it would do if the near and far views were taken by a camera. When we watch a person moving about a room he does not seem to alter materially in size. Yet we know that his image on our retina must change considerably when he steps a few feet away.

The perspective of interest is a name which we give to certain tendencies to distortion. Some of these are shown in

[6] G. Adams, An experimental study of memory color and related phenomena, *Amer. J. Psychol.*, 1923, **34**, pp. 359–407.

the visual arts, especially in primitive representations of human and animal forms. The tendency is to exaggerate the size of the head in comparison with the body as a whole. There is also an exaggeration of the face as compared with the

FIG. 2.

head as a whole, and an enlargement of the eye with respect to the face. Idols from many parts of the world show these distortions. So do early Greek statues, Byzantine paintings, German wood carvings, ornaments from South America, drawings by North American Indians, etc., etc. Figure 2

illustrates the point with images from widely scattered sources. The same tendencies are often found in sketches by children, and also among untrained adults. Recently I asked one hundred students to draw from memory (1) a profile of a human head, (2) a human head front view and (3) a full length human form front view. Most of these persons were untrained in drawing. Only a few drew their figures in approximately correct proportions, and the others distorted their figures, invariably in the same direction, namely, by enlarging the head. The question must be asked whether the artist actually perceives his models in this distorted form, or whether the distortion is merely in his execution of his idea. In the latter case the distortion may be intentional or unintentional. Experiments could be devised to test whether there is a genuine 'head-size illusion.' They would be a welcome contribution to experimental æsthetics.

The last few paragraphs have been devoted to naming some of the sources of the individual views of objects which we call perspectives. There is a sense in which every perceptual moment is unique, and these perceptual differences must be recognized as one important source of the imaginative differences between person and person.

We come now to the question of the relations among these perspectives. If each moment of our lives is unique, each percept individual, how do we refer certain groups of percepts to the same object? The question has two forms, one individual and the other social. The individual form of the question asks how I can know, that the same object is revealed through so many different percepts. Why, for example, do I attribute to the same cat such contrary qualities as her soft fur and her sharp claws, her loud purr and her silent tread? The answer is that the cat really combines those qualities, and that I also combine the impressions of them through my past encounters with the cat. It is I who supply the unity by which the cat's unity is known to me. The continuity of the individual subject is a necessary background against which the continuities of objects may be realized. That is, the recognition of any object implies not only that the object was there

before, but that the person who recognizes it was there also. Recognition, then, is the attribution of a common past to oneself and the object.

Philosophers warn us against the idea that our percepts are copies of the objects to which they refer, but there is a certain sense in which the percept may truly be a likeness. What I mean is that a present percept is always reminiscent of a previous one. It always involves some degree of recognition and identification. Even the most novel object looks like something, or sounds like something, or feels or tastes or smells like something. It cannot be a percept and be totally incapable of classification or assimilation to other parts of our life. Whatever is grasped at all is grasped as something else than just itself. In this sense it is indeed a likeness. The question how a single object is constituted out of a variety of percepts is a special case of the problem of association. Certain elements out of the present scene become assimilated in peculiar ways to other elements out of the past, and thus objects are evolved. It is in the theory of memory if anywhere that we may look for explanations of perceptual objects.

The question of the integration of perspectives has also a social form, namely, can two minds meet and communicate about the same object? Can two persons perceive the same thing? There are countless apparent examples of it. I watched a two year old baby and a young bull pup each seize a corner of a handkerchief which had been dropped on the floor. They pulled in opposite directions and seemed to be in direct communication with each other. The baby gave a pull and the dog answered with a jerk. They then settled down for a tug of war. Neither of them had any articulate language, but they appeared to understand each other. Moreover a spectator could see that they had hold of the same object and that they meant to hang on. It is clear that such an example brings us back to the question of the perceptibles. Can a spectator really 'see' what the contestants mean to do?

In a previous paragraph it was said that the only things which a sense organ can give us with complete certainty are the strictly sensuous elements. And this I believe remains

true. But I wish now to emphasize the fact that for practical purposes we must assume that the senses do give us more than these sense elements. In accepting these objects as given we run the risk of illusion. But unless we take this risk we must pay a penalty for our caution. If we waited for complete certainty with respect to sense objects we should be paralyzed by our own skepticism. The boxer, the fencer, the tennis player must all be able to perceive not only what their opponents *are* doing, but also what they are *going* to do. And although there are times when the player's guess is wrong, this is better than not to guess at all. Likewise an automobile driver making his way through heavy traffic must react to what other drivers are going to do, and to what pedestrians are going to do. We may conclude that there is indeed a transcendence of individual perspectives in the case of perception, and that it is an observed fact that two persons react to the same object.

Such a combination of perspectives depends upon the associated activities of different persons. The concept of 'the same object' as viewed by different persons would not arise unless there were coördinated acts as between these persons. Common purposes cause us to perceive common things. These purposes may be competitive, or they may be coöperative. When two children fight for the same plaything the acts of each child help to define and accentuate the meaning of that plaything for the other child. In a football game, although each spectator has a different view, all assume that they are looking at the same ball. Different perspectives sometimes lead to different beliefs as to what has happened to the ball. When these differences have to be settled we have a referee or umpire to settle them, and what he says 'goes' or the sport could not continue. The referee's decision is the official truth. It is safe to say that social activity could not continue without such official or pragmatic perceptions.

Many persons have pointed to the fact that our motor responses affect our percepts. The best statement of such a motor theory of perception which I have ever read is that of Washburn [7] who, in her thesis, includes the statement that "All

[7] M. Washburn, Movement and mental imagery, Boston, Houghton, Mifflin, 1916.

association is the association of movements." And the bearing of this upon perception is stated as follows:

> In first making acquaintance with an object we respond to it as an undifferentiated whole: later we come to make specialized responses to various parts and aspects of it; but it is the fact that it can still be responded to as a whole that keeps these specialized movements together in a single system, and thus gives the object its unity. An orange, or a chair, or a tree, is a single object, and not a mere aggregate of qualities and parts, because it can be reacted to as a whole, and because every one of the movements involved in attending to its parts is associated with the movement of reacting to the whole object. Some of these movements which are associated with various parts or aspects of an object are compatible: the fact that they lead to the common outlet of a movement made to the whole object may thus transform them into simultaneous systems. Others are incompatible, and must be united in successive rather than simultaneous systems; thus, for instance, one cannot attend simultaneously to both ends of a pen, but the whole object is a unity because of the possibility of responding to the whole of it by a single movement or movement system.

It is possible to extend Washburn's theory to include the social group and its apprehension of a common object. A drama is a very good example of a collective object. It is produced collectively and perceived collectively. The idea is expressed in one of the essays of Miguel de Unamuno,[8] where he writes:

> The theatre is something collective; it is where the public intervenes more and the poet less. At one time, the time of youthful vigor, the conditions of publication were very different from those of today; a drama remained in manuscript much more than today and very much more subject to continual revision and improvement, by those who gave suggestions at each new representation.

In saying that a dramatic performance is perceived collectively I mean that the individual spectator is influenced by the presence and the reaction of the audience of which he is a member, and also that the performance of the players themselves is changed by the temper of the audience. Plays are

[8] La regeneración del teatro español, 1896, in Ensayos, por M. de Unamuno, Pub. de la Residencia de Estudiantes, Madrid, 1916.

literally made by the reactions of audiences, directing the effort of the playwright.

Other examples of perceptions which have a public character are observations of scientific research workers. Here the situation is different from that of the audience, because the scientific observer takes pains not to be influenced by others at the time of his observations. But in planning his experiments, and in communicating his results he assumes that other observers, if they follow his directions and repeat his experiments will perceive the same things that he reports. If it turns out that they fail to verify his observations, then the science must await still more observations on that subject. The work of science is social work, and we may say that the perceptions of the scientist are affected with a public interest. The individual is the instrument through which the group gathers perceptions. His observations are made possible not only by material support from the group but by the system of ideas and the instruments which constitute the tradition in the particular field which he studies. Generally speaking the individual will not spend his life on the type of problem which has no prestige with his group. Our perceptions are influenced by the ideology of the group because that is what directs our attention to one field of observation rather than another.

Summary

Things perceived and things imagined shade gradually into one another. We have classed perceptual images as imaginative constructs because (1) such images reflect individual perspectives, perspectives due to the individual's structure, history and situation, and (2) the content of the individual's perceptions is due in part to social selection. In trying to enumerate classes of perceptibles we have included simple sense qualities at one extreme and have indicated that at the other extreme we could perceive some cases of human emotion, character and action, and some features of abstract ideas.

[MS. received December 5, 1934]

EXPRESSION IN MUSIC: A DISCUSSION OF EXPERIMENTAL STUDIES AND THEORIES

BY KATE HEVNER

University of Minnesota

In the study of æsthetics from the psychological point of view, one of the most important problems is the meaning or expressiveness of certain lines, forms, colors, rhythms and sounds. By the expressiveness of these elements, we mean the immediate and direct responses which the organism makes to them, especially such physiological responses as motor attitudes, empathic experiences, affective states, mood effects and emotional reactions, since the sentiments or concepts or ideas which may also be aroused are largely conditioned by these physiological responses. It is the purpose of this study to outline the various aspects of the problem of expressiveness in the field of music, to discuss the different theories which have been proposed toward the solution of the problem and to evaluate some of the experimental material which has accumulated around it.

The Nature of the Problem

It will be necessary at the very outset to have a clear understanding of the limitations of expressiveness in music, in order to evaluate some of the apparent controversies which have clung to the problem. There has probably never been much argument over the presence or absence of some kind of meaning in music. Both historical and experimental results give evidence of its existence. We have, for example, the important and well understood use of symbolic music in the modern opera, and in the tone poems and other forms of descriptive music which have met with outstanding popular success. The numerous volumes of program notes also testify to the popularity of symbolism with concert audiences. In experimental psychology there is evidence that the general

mood or affective state suggested by any one composition is often fairly constant and universal.[1] The introspective reports of individuals who try to analyze their experience in listening also show that this interpretative experience is of very frequent occurrence,[2] and furthermore that it is often the means by which apprehension or appreciation of a more difficult type is developed. On the other hand it is a very simple matter to demonstrate, under different kinds of experimental conditions, that listeners often disagree very markedly in their interpretations of music. Some of the differences of opinion of eminent critics on important symphonies have become world famous.

There is probably no one who would maintain either that music invariably has a meaning, that all listeners give it meaning, that any one listener always gives music meaning, or that there is always agreement on the meaning of any one composition. Likewise there is probably no one who would maintain that the opposites of these statements are true,—that no music has meaning or that there is never agreement on the meaning of any one composition. In fact, the controversies which have raged over this problem all start from the common ground that there are individual differences both in human listeners and in musical compositions. The form and structure of the music is important in determining whether or not an interpretative attitude on the part of the hearers is called for, whether the meaning is clear-cut, easily comprehended, and generally agreed on, or whether it is obscure and difficult and fluctuating. The momentary moods and the personalities of the listeners also determine the expressiveness of the music. Their attitude and their success in apprehending the music vary with their training and sophistication in musical matters, their present affective states, their physiological well-being and their backgrounds of musical and non-musical experience. A perfect agreement, an absolute universality and a rigid symbolism are most certainly disclaimed. In fact a large

[1] M. Schoen, (Editor), The effects of music: London, 1927, Ch. VII, VIII.

[2] M. Schoen, The æsthetic attitude in music, *Psychol. Monog.* 1928, 39, pp. 162–183.

flexibility of the meaning of any one composition is not only postulated but assumes a certain importance as a source of æsthetic enjoyment of music. It is due to this flexibility that each individual listener may find in the music just that particular shading of its general mood which is most desired and best suited to satisfy his individual demands. There is nothing to prevent him from applying the general characterization of the music to the specific personal experience which he wishes to associate with it, and he is encouraged to assume that the mood of the music, stated in these more general terms, is ideally adapted for the expression of the specific personal experience with which he is led to associate it.

This same flexibility, however, is probably the reason for the opposing points of view which have sprung up over questions such as the following: Is the meaning widely and generally understood by all kinds of listeners or apprehended by only a small percentage of listeners, supposedly those trained in music? Is the meaning only a very general suggestiveness of mood or feeling, or is it exact and explicit, arousing definite emotions, sentiments and ideas? Is there or is there not a system of expressiveness, with elements of musical structure, such as rhythms or chords, to which certain ideas or emotions are always attached, and which can be repeated and combined by the composer with constant and predictable results.

These questions are of course interrelated, and two of them, at least, may be considered together. How definite the meaning is, and how widely it is understood, may be restated in more quantitative terms: Is there one modal reaction to any one musical composition and does it occur with high frequency in unselected groups with little deviation from the mode? Experimental attempts to answer these two questions have brought out apparently conflicting evidence, briefly, as follows: (1) When long musical compositions are heard, such as movements from symphonies, preludes from operas, etc., and when the listeners are asked to report their reactions, either by writing introspections or by matching the compositions with a stated list of emotions or moods, the reports always show a

variety of meanings for any one composition. (2) Compositions in the minor mode are often interpreted as gay,[3] and certain pieces in which the intention of the composer was obvious, such as the Hymn to Joy from Beethoven's Ninth Symphony are given interpretations rather far afield.[4] (3) When isolated elements are presented, as for example, major chords vs. minor chords, differences in the affective tone are not perceived except by the highly trained.

These results, however, which would seem to indicate that musical meanings were utterly chaotic, may perhaps be accounted for by certain sources of error in the methods, and a critical examination of these methods may clear the way for a more orderly system of musical expressiveness. In the first place, the experiments often do not take into account the quantitative differences in the reports, so that the large numbers of reports which are in general agreement with the composer's intention are not sufficiently emphasized. The one report of 'despair' for the Golliwog's Cake Walk acquires as much importance as the dozens of more suitable responses. On the other hand, if each subject is allowed to choose as freely and as frequently as he likes, from a long list of adjectives, arranged into groups for his convenience, the experimental evidence is overwhelmingly consistent, and the occasional discrepancies may be viewed in their true proportions.[5]

Among longer compositions such as are often used in studies of meaning, very careful choices must be made by the experimenter in order to include only those compositions in which one mood or affective state is maintained throughout the entire structure. Very often the mood of the composition is changed in the middle section, a useful device whereby the original mood, when it re-appears is intensified by the contrast effect. Nevertheless, the listener whose attention has been caught by the contrasting section will report a response not in harmony with the dominant tone of the composition as a whole. Recent experiments give very clear evidence of the

[3] J. E. Downey, A musical experiment, *Amer. J. Psychol.*, 1897, 9, pp. 63–69.

[4] D. Moore, Listening to music: N. Y., W. W. Norton & Co., 1932, p. 11.

[5] K. Hevner, The affective character of the major and minor modes in music, *Amer. J. Psychol.*, 1935, 47, 103–118.

importance of this factor. For a piano piece of Debussy, 'Reflections on the Water,' no consistent pattern of meaning was discerned, and the reports of the listeners were scattered and conflicting. However, when the three sections of the piece were separated by the experimenter in the presentation, each section revealed a very definite pattern of response, although in quite different parts of the affective scale.[6]

Much of the apparent disagreement among listeners over the meaning of music could be avoided by an experimental method in which two certain factors were at a minimum,—the slightly different connotations which individuals place on specific words, and the effect of the immediate circumstances under which each individual hears the music, *i.e.* his physiological condition, temperament, recent experience, etc. The method of grouping together several words descriptive of the same general mood so that the listener has some choice in voicing his responses, and so that the experimenter by combining these individual choices into the larger groups, may eliminate some of the chance or irrelevant errors, seems to be a step in the right direction since it permits the more generalized mood effects to emerge and at the same time allows a quantitative treatment of the data. No one word has an absolute meaning which could be agreed on by one hundred per cent of the listeners. The differences may be very small, but they are very real. They may be important to the individual listeners, but they are not of importance to the experimenter. The momentary mood of one listener may give sad music a doleful aspect, while the mood of another may cause it to appear tragic. Thus the same music may seem to one individual agitated, to another impetuous, to another exciting, to another both exciting and triumphant, but the meanings of all of these words indicate a consistency and unanimity of response, and the experimenter should make the most of that truth by combining all of these words into one group and treating them as an entity. It is not of any great importance, for example, to find that pathetic and melancholy are slightly more significant

[6] K. Hevner, Experimental studies in the expressiveness of music. To be published, *Amer. J. Psychol.*, 1935.

than doleful and depressing in differentiating the character of the minor mode.[7] These differences may be due to the general attractiveness and usefulness of the words, or perchance to the position of the adjectives as they appeared on the page. It *is* of importance to discover that a group of ten adjectives denoting various aspects of sadness and a similar group of adjectives denoting joy, will differentiate very sharply the minor from the major mode.

If we return then to the question: Is there a clearly apprehended meaning in music, one definitely limited response which occurs with high frequency, and with little deviation, the best answer seems to be as follows: (1) studies with adjective groups covering eight well defined affective states or moods, and with many varieties of music, long and short examples, in major and minor, and with different rhythms, tempos, and harmonic structures, indicate a surprising uniformity and consistency in the apprehension of musical meaning, and (2) there is no important and irrefutable evidence to the contrary.[8] In regard to the further question of how widespread is this understanding, the best answer seems to be: the meaning of music is one and the same for listeners of all kinds, the intelligent and the unintelligent, the trained and the untrained, although those who have the most training differentiate more sharply and with greater ease. This conclusion is based on two lines of evidence which supplement each other very nicely. (1) Studies on the expressiveness of music by the adjective group method and with many varieties of music have included the reports of more than 450 different persons (college students) who, fortunately, had been measured for general intelligence, appreciation of music, musical interest, talent and training. Especially in their formal music training do we find large differences. The range includes those who have studied many years and intend to make a profession of music as well as those who have never taken a lesson nor played a note, who never attend concerts, and never turn on

[7] *Ibid.*

[8] We have already pointed out certain discrepancies in method and for some further explanations see C. C. Pratt, The meaning of music, McGraw Hill, 1931, p. 211.

the radio. But the differences between the untrained and highly trained groups are but slight, and when studies are made of the poorest groups, the same responses stand out, although a bit less distinctly.[9] (2) Differences between the trained and the untrained may, however, be brought out by different methods.[10] If the minor qualities for example are confined to single chords, without the additional accumulation of stimuli which would result from twelve to sixteen measures played in the same mode, then the apprehension is made so difficult that only those with most training can achieve it, and the affective state which goes with the mode occurs only indirectly, as a sort of afterthought, entirely dependent on the identification of the modality itself.

The Sources of Expressiveness in Music

In reaching the conclusion that an intelligent interpretation of music is widespread, we do not specify the exact source of the associations which constitute the meaning or the expressiveness of the music. We define meaning simply as a way of responding to an objective stimulus, and when we refer to the character of a composition we have reference to the most frequent response of the greatest majority of listeners. To make the claim that there is a substantial agreement in regard to the existence of a characteristic response, implies only that it is a trait found generally among the human beings we have tested, not that it is a native and inherited or an acquired and cultivated response. Our present experiments give us no data with which to deal with the question of the source of the associations. That constitutes another and a separate problem, which cannot be solved without empirical data from further experiments, and from experiments which, we might point out, would be particularly difficult to set up. Our social and cultural inheritance is so important and effective in building up our habits of perception and apprehension that it is impossible to conceive of experimental conditions

[9] K. Hevner, *Op. cit.*

[10] C. P. Heinlein, The affective character of the major and minor modes in music, *J. Comp. Psychol.*, 1928, 8, pp. 101–141.

which would control them. In the absence of experimental evidence bearing directly on the explanation of meaning in music, a satisfactory theory may be derived indirectly from evidence contributed by experiments on related topics.

The associations between affective states or attitudes and certain musical experiences must take place in the same way as other associations, such as those between sounds and concepts in language and between the stimuli and responses in any learning process. They are strengthened according to the laws of frequency, of effect, of vividness, practice, redintegration, etc. We will find that the similarities between musical sounds and affective or emotional states, that is to say, the ties which bind them together, may be of several kinds. Perhaps the earliest were physiological states, or bodily activities, the quickening of the pulse while listening to a strong regularly accented beat and while marching along in a patriotic parade; the relaxation under soothing warmth of the summer sun and the relaxation from slow rhythms and smooth changes in harmonic progressions; the incipient movements in watching a dancing figure and in following a turn or grace note in a melody. Or the associations may spring from a similarity in visual or tonal patterns: the careless playfulness of a descending mountain stream may be carried over to a quickly tripping descent in a melodic pattern; the same feeling of supernatural force and fury which we get from the fitful changes in intensities during a violent thunder storm may be associated also with the irregularities of intensity, tempo and rhythm in music. Finally a frequently repeated meaning whether logically appropriate or otherwise, may, by the force of its popularity, stamp its meaning on all similar stimuli; for example, the low register, monotonous repetition and limited range of melody in the well-known 'Miserere' from *Il Trovatore* may affect the future expression of tragedy in music, or the short regular phrases, and $\frac{4}{4}$ rhythm of 'Onward Christian Soldiers' may stamp the pattern for the expression of religious enthusiasm in this western musical world.

The Use of Expressiveness in Understanding and Appreciating Music

We find it quite possible to understand why human nature has found it so easy to fall into the habit of listening to music in this interpretative way. Even though it may be a purely acquired attitude, dating from the oft repeated lessons in the first grade,—"What does the music say?"—it obviously meets with no serious opposition. Students asked to find meaning in musical compositions exhibit the greatest docility and ease in complying with the experimenter's request. It is reasonable also to expect that it will be the easy and pleasant method in contrast to the more difficult method of organizing and apprehending musical experience which might call for more abstract symbols and for concepts which our present day music training does not provide. For example, it is easier for the average man to follow a certain musical composition by the first than by the second of the following descriptions:

(*a*) We see the shadows of the pines which overhang the entrance to a catacomb. From the depths rises a chant which reëchoes solemnly, sonorously like a hymn sung by some furtive Christian brotherhood and is then mysteriously silenced. . . . Then, indistinctly, incessantly the rhythm of innumerable steps. There is a vision of past glories, trumpets blare, and the army of the consul advances brilliantly in the grandeur of a newly risen sun, mounting in triumph the Capitoline Hill.

(*b*) The first subject (*piano e sotto voce*, F minor, $\frac{2}{2}$ time) is a swift line, drawn for some distance without harmonic perspective by the strings and bassoons. Somewhat varied in rhythm it is repeated in the winds with the arpeggio accompaniment in lower strings. It leads to a second subject characterized by a triplet figure in horns and celli with continuation in the winds, and goes on into a march like closing subject syncopated and vigorous. Development begins with the opening subject in the winds and the last half of this theme, augmented and in sustained notes forms a *canto fermo* for a polyphonic development which culminates in the long expansion of the sober episodic phrase above noted.

Both descriptions enable the listener to follow a composition and recognize and identify its various parts. They are

helpful in keeping his attention on the music, and in comprehending its structure, its various parts, and the appropriateness of its parts to each other and to the whole. But the first is for most people an easy and pleasant way of accomplishing these ends, and the second is more difficult and uninteresting. The first does not presuppose any concepts or experience except those which are the common property of all human beings, pine trees, singing, armies, triumphs, etc. The second presupposes a larger and more specialized vocabulary and a wide experience and training in music. The first deals only with the most realistic and concrete terms, while the second is more abstract and technical. The first is made attractive and colorful, suggesting pleasantly toned affective states. The second calls not only for finer sense discriminations but also for a greater amount of voluntary attention in manipulating these details, and both of these are intellectual activities which are likely to be unpleasantly toned. Small wonder that the first type is the more popular with our present day audiences. Listeners have learned to like the progammatic music better because they can find in it something concrete and tangible, which gives them the key to its structure. Music which does not lend itself readily to such analyses they reject as uninteresting and dull. Small wonder also that the skilful musician or sophisticated listener for whom the intellectual activity in comprehending music has been perfected through much practice, resents either type of exposition as an unwarranted intrusion comparable to that of the rhetorician who might explain the sentence structure and the prosody of a page from Shakespeare.

The Nature of the Symbolism in Music

We have already made clear our belief that the expressiveness of music is of a general rather than a specific nature, arousing an affective state rather than one specific emotional experience. We have also explained that this quality in the expressiveness of music has importance in producing a pleasant, and an æsthetic experience in the listener. Our own opinion is based on experimentation directed toward this

particular question, although other students of the problem have arrived at similar conclusions by a number of different routes. Pratt [11] has explained that the function of music in arousing the genuine emotional state, as psychology defines it, is quite negligible, and that those who maintain that music is the language of the emotions, use the word emotions for a state much less explicit and well defined, a state in which, to use the language of James, the bodily sounding board is at work. "Intense enjoyment and pleasure are frequently referred to, in popular speech as rapture, elation, delight, transport, exultation, ecstasy, etc. whereas in psychology these names would more generally be regarded as variations on the single theme of intensive pleasantness." [12] Ferguson [13] brings out still another point of view, namely, that music produces the bodily reaction, or the organic background of the reaction, directly, and leaves the interpretation, the introduction of an appropriate external cause, to the individual. This is in agreement with some modern explanations of emotional experience which point out that the bodily reactions for certain classes of emotions such as rage, anger, fear, hate are similar or even identical, but that the differentiation of these different emotions one from the other, depends on the external stimulus.

There is another characteristic of the symbolism of music which is also of importance and which has been implied in our previous paragraphs, although not explicitly stated: the symbolism of music is based on a complex but orderly system. To claim that certain compositions have a constant and generally understood character might mean merely that these pieces were entities of expressiveness, with a meaning which emerged only from the composition as a whole and in which the variation of a single part could have no predictable effect but might bring about a complete metamorphosis or nullification of the previous effectiveness. We believe, however, that by modifying certain structural elements in the music, by increasing or decreasing the loudness, by shifting the rhythm

[11] C. C. Pratt, *op. cit.*, pp. 160–161 and *ff.*

[12] *Ibid.*, p. 177.

[13] D. N. Ferguson, On the elements of expression in music, Papers and Proceedings of the Music Teachers' National Association, 1925.

or raising the pitch, we can modify the expressiveness of the piece and bring about certain slight but predictable changes in the mood effects. That we can make these modifications in the structure of the music and predict, or at least learn to predict their effects, implies that there are elements of expressiveness which may be manipulated, modified and varied, and whose attributes are relatively constant and fixed. The rhythm of the composition, for example, has an expressiveness of its own, and contributes a tone of sadness, or of lightness and gaiety, which may be modified by varying this one factor alone. Likewise, a slow tempo usually carries with it the feeling tone of languor and relaxation, and, if not obliterated or overbalanced by other elements in the musical structure, may be expected to suggest this character, and when combined with various rhythms or harmonies, may express a variety of related moods, sadness, tranquillity, dignity, etc.

The claim that the expressiveness of music is based on a complex but orderly system is not made with the conviction that the system may be discovered directly by a series of crucial experiments, or that there will never be exceptions to such principles as may be evolved. It is made with full consciousness of two very great difficulties: (1) the many sources of error, false steps, and contradictory evidence which always obstruct the progress of the scientific method in spite of the heroic efforts of experimenters to eliminate them, and (2) the very closely knit and subtle interrelations of all the parts of any truly musical composition. This second difficulty is of special importance because of the method which we advocate for the study of this system, a method in which several versions must be devised for each composition used, versions which differ from one another in one respect only, in the rhythm, or tempo, or harmony, etc.[14] It is difficult to

[14] This method of altering the music is not as barbaric as it might seem at first thought, since it is very little different from the method of composers in writing a theme with variations. It is very difficult to find music which will lend itself to this kind of treatment, and exceedingly difficult to write the alternative versions. The services of the most expert musicians are required for both processes. The new versions are perhaps not always of the same high standard of excellence as the originals although on the average the subjects like them as well, and sometimes even prefer

alter one factor in a composition without mutilating the composition and reducing (or extinguishing) its musical qualities. A certain melodic line demands a certain harmonic support; one cannot be varied without the other. Or if the harmony is arbitrarily changed, the melody will be perceived as a different melody even though the actual sequence of tones remains unaltered.

The understanding of musical expressiveness, and the search for a system of symbolism in music, has been confused and delayed by mistakes in choosing the units, which carry the meaning. In selecting the structural units, or carriers of meaning, we must certainly consider the factor which distinguishes music from the visual arts, its dependence on temporal rather than spatial relationships. The structural units which promise the most success will be not those which occur at any one moment of time, but those which involve successions of stimuli, whose very existence depends on what has gone before and what comes after. To the layman, one of the most obvious units of musical structure is the chord, just as the word is the most obvious unit of language. But the chord, when a lone and single stimulus, has not proved satisfactory as an element of musical expressiveness. Even when we are dealing with as familiar and generally understood a symbolism as the sad quality of the minor mode, we cannot establish the sadness of any one minor chord appearing alone.[15] We are too much in the habit of perceiving music in terms of time relationships. Our concept of the form, the unity, the balance, of any musical structure, our very perception of music as music, is based on these successions of stimuli in time.

Some of the most interesting contributions to the problem of meaning in music have come from the analyses of musical material of one type, such as sad and tragic music, love songs, martial music, etc., in the search for the forms and structures characteristic of the type. Only the expert musician is qualified to make such contributions, and two noteworthy

them! Neither do they in all cases represent a change of one element only, although every attempt is made to satisfy this important criterion. Fortunately, scientific truths are sometimes discovered under slightly less than perfect conditions.

[15] C. P. Heinlein, *op. cit.*

analyses of this kind have been made. Gundlach [16] studied a large number of Indian songs, which he divided into three groups, healing songs, love songs and war songs. He found that the most striking differences between the groups were in the rhythms, a fact which might be interpreted as giving rhythm the place of first importance in expressiveness. The second study, which deals with music of more interest and importance has been made by Sorantin,[17] who has studied four different expressive forms—joy, grief, longing, and love.

The value of Sorantin's method of procedure depends on the representativeness of the author's sampling, and on the evidence that his choices are true samples of the mood in question. In both of these respects, it seems to this reviewer, some errors have been made. There are twenty-one examples for joy, taken from operas and songs where the intention of the composer was obvious, but the figures of joy are also based on sixteen examples from 'pure' music, including symphonies, sonatas and concertos where the intention of the composer is more difficult to identify and the import of the music somewhat questionable. The experimenter's contention that 'the so-called absolute or pure music, especially in the case of Bach seems to be far more expressive of highly intensified emotions such as despair than music composed for a given text,' [18] is not particularly convincing nor is his support of his thesis by the analysis of three of his own published compositions of absolute music.

To some musicians and composers who have studied the results of Mr. Sorantin's findings, the many examples of expressive music which do not fit his conclusions loom up with a great deal of importance. On the other hand, some of his suggestions including the patterns of simple consonant harmonies and circular rhythms for joy, and dissonance and slow tempos for grief seem to meet with more approbation and enthusiasm.

[16] R. H. Gundlach, A quantitative analysis of Indian music, *Amer. J. Psychol.*, 1932, 44, pp. 133–145.

[17] E. Sorantin, The problem of meaning in music, Nashville, Tenn., Marshall and Bruce, 1932.

[18] *Ibid.*, p. 56.

Perhaps some of the peculiarities of musical symbolism may be brought out more clearly by comparing and contrasting it with verbal symbolism. The symbolism of language is not only much more generally understood, and more fixed and rigid than that of music, but it is much less dependent on relationships, temporal or spatial, for its meanings. To be sure, the meaning of certain words may vary to some extent with the content, or with the past experience of the reader, but there is a specificity and independence in their meanings, which is not to be found in music. The elements of language may be combined and manipulated in many ways without losing their identity or being affected by the juxtapositions. For example:

The child	screamed	lightly
The President	jumped	aimlessly
The dog	ate	eagerly

are all separate ideas which may be put together in many ways, with little or no modification or fluctuation in their original meaning. On the other hand, musical elements such as tones or chords have no fixed symbolism of their own, and acquire a symbolism only in continuity, in relation to each other. It is in their juxtaposition alone that the meaning or character is acquired. There is no dictionary of musical language where one may look up the chord for 'apple,' the pitch for 'river,' the rhythm that means 'good fortune,' and we repudiate any system of explaining the meaning and character of music which implies a fixed symbolism for certain specific sounds, *i.e.* for certain melodic skips—the upward skip of a fourth, the downward skip of an octave, etc., or for certain chords—the tonic triad, or the dominant seventh. We do not believe that the use of ascending melodic skips of sevenths will introduce the idea of longing, and if combined with sixths and thirds in the harmony will be interpeted as love-longing, tending toward desire, and that a flowing rhythm which is more feminine than masculine will identify the feeling as that of a woman, nor that the circumstance of a low register and an end phrase of dominant seventh resolved to tonic will indicate that the heroine has been unhappy in her love but has become at the

end, resigned. Any such system of meaning in music would be ridiculous on three counts: (1) music produces only general mood effects and is the stimulus only for the bodily sounding board, not for specific ideas; (2) musical elements such as melodic skips, resolutions, rhythms, are not heard as entities and they cannot maintain a specific and fixed ideational concept because they are inextricably woven into the fabric of the music and are modified and even metamorphosed by their influences on each other; (3) since music is a temporal art, its structure and effectiveness depending on its successions of stimuli, the elements of musical expressiveness must be those of motion and stress, involving repetition and relationship, sequences and successions.

We do believe that a swift tempo combined with a lively rhythm, simple major harmonies, and light intensity in a high register will produce a mood effect of gayety and exhilaration, and that the mood can be inclined toward agitation by changing the rhythm, or toward the eerie and strange by altering the mode, toward the playful and delicate by changing the intensity, toward serenity by lowering the register or modifying the rhythm, etc. . . . Unfortunately, we cannot predict the exact changes in these formal elements which will be necessary to change the mood effect, because we have not accumulated sufficient experimental evidence, but we believe that it is logical and profitable to proceed with experiments along these lines and that properly controlled experiments will yield consistent and orderly results.

To make this claim is quite different from making the claim that the major mode is always happy and gay, or that any mood effect is always associated with some one formal element. There are several means by which music may be made happy and gay, and a harmonic effect, such as modality is only one, although it happens to be a very good one. The effect of this one factor may be overshadowed or obliterated by the effect coming from the rhythm or tempo, etc. . . . In the face of the very merry scherzo of Mendelssohn in G minor, it would be ridiculous to maintain that the minor is always sad. But we do maintain that the general affective tone of the minor is

sadness, and of two compositions with similar rhythms, harmonies, melodies and tempos, the one written in the major mode will always be *more* merry and gay. The way to prove it in this case would be to rewrite the Mendelssohn Scherzo in the major mode, and we predict that it will be a little merrier in the major than it is in the minor. It may lose some of its quaintness, its strange, fleeting remoteness or some other quality which we consider the essence of its charm, but the sheer brightness and joyousness of the piece will be enhanced, by the change in the mode.

The theory or system for musical expressiveness which seems to be at the same time the most logical and the most profitable for experimental work is the system suggested and used by Mr. Ferguson[19] in his analyses of the problem of meaning in music. For him there are two kinds of elements basic for all music, 'motion suggestion' and 'stress suggestion.' The first of these carried by the tempo and rhythm is the tendency toward movement, the impulse to follow the activity of the music at its own speed and through its accented repetition by means of the direct bodily and motor responses which are set in motion by the successions of sounds. This suggestion of motion is one of the most powerful effects of music. It simply compels our attention and absorbs all of our motor activities. It pulls us into line and determines the direction of our forces. We must move, and move with the music. The 'stress suggestion,' as the term implies is a matter of intensities and extensities of response, of tensions and relaxations, of efforts and restraints. Both of these elements are strongly directed toward the bodily sounding board, bringing about physiological changes, in the pulse, breathing, circulation, muscle tension (with attendant kinæsthesis) and probably visceral and glandular responses which are difficult to trace. The stress elements may be sub-divided into three kinds, two in melody and one in harmony. In melody there are: (1) the rising and falling of the pitch, and (2) the relation to the tonic or keynote. We might describe the first of these as the movement pattern of the melodic line, its climbing or falling, its

[19] D. N. Ferguson, *op. cit.*

circular motion.[20] The second, we may explain as the excursions of the melody through and around its central point or keynote, involving the difference between the rest tones, or closely related tones, (1, 3, 5) and active or more foreign tones (all others in the scale). In harmony there is consonance and dissonance or smoothness and harshness, in all its variations of position, quality and degree, and the interrelationships between successive harmonies. All these elements may be used in as many permutations and combinations as musical taste will allow, *e.g.* active melodic tones harmonized with the harshest discord and placed in the strongest rhythmic positions, in contrast to rest tones, harmonized by pure consonance and without rhythmic emphasis. There are also a number of attributes such as the tone quality of different instruments, register, key, and various dynamic effects all of which modify or enhance the expressiveness although they are merely adjuncts and not elemental in themselves.

By means of experiments aimed directly at these elements and certain combinations of these elements, we believe it may be possible to discover at least the main outlines of a system of expressiveness in music. We may be able to discover which are the most important and stable characteristics of the various elements, which the most modifiable and fluctuating, and how the various elements enhance or modify each other. Our experiments, based on this system, have already given some indications that the rhythmic and harmonic elements (including modality) are of stronger effectiveness than the movement of the melodic line. Other principles, especially those concerning the interrelations of the various elements, may be expected to emerge from the application of variations of our methods.

These discoveries which will be made through these indirect and clumsy applications of several varieties of the 'scientific method,' can then be put forward as *new* principles, although they have been in existence for many centuries, have been practiced extensively by all successful composers, and have been handed on from one generation of composers to another

[20] Cf. Sorantin, *op. cit.*, 'dolciata,' p. 91.

by (they would probably say) intuition or instinct. But our 'facts' will be more useful in understanding for ourselves the excellencies of good music and in teaching and explaining it to larger and larger audiences. If the great artist could speak to his audience verbally as effectively as he does musically, our efforts would be unnecessary, but he seldom expresses himself except through the medium of his art, and when he does, it is usually not in the terms calculated to be most useful and helpful.

[MS. received November 5, 1934]

DRIVE AND THE CHARACTERISTICS OF DRIVEN BEHAVIOR

BY MERLE HUGH ELLIOTT

Harvard University

Drive and the various related motivational concepts are coming to occupy an increasing amount of space in the psychological literature, and as subjects for polemics they seem destined to take the place formerly held by 'instinct.' Criticisms of such concepts have arisen naturally because of the loose and varied usages of the names applied to them and the many consequent misunderstandings. Other criticisms have developed as a result of the mechanistic suspicion that such concepts are essentially teleological.

The present paper takes the point of view that drive is nonetheless a fruitful concept which can and must be given objective definition. After first defining the term it attempts to derive from the available experimental material some general principles regarding the objective characteristics of behavior under drive and criticizes, in the light of the principles developed, some of the experiments which have purported to measure 'strength of drive.'

Definition of Drive

In order to be useful to the experimentalist a term must be definitely related to the actual variables which can be manipulated and controlled in the experimental situation. It will be declared at once that in experimentation drive most usefully refers to *the total of internal changes produced in an organism by some kind of deprivation.* Thus the organism which has been deprived of food for a certain time may be said to act under the influence of the hunger-drive. The strength of the drive is controlled experimentally by the extent of the deprivation, in the case of the hunger-drive by either the length of the fast or by the amount of deficiency of the ration.

This treatment of drive does little violence to established usage and has the merit of clarifying the concept. In motivational experiments the independent variable is deprivation. The dependent variables are various aspects of behavior. The motivational experiment is primarily concerned with correlations between deprivation and behavior. Drive is thus a construct. It is an intervening entity which is set up to assist in explaining correlations between deprivation and behavior. Each drive may be specified eventually, it is to be hoped, in physiological or biochemical terms.[1] For the present, however, drive must remain as a more or less hypothetical resultant of deprivation.

It may be objected that the concept of drive is superfluous, that the terminology of deprivation and behavior, or of stimulus and response, is ample for the treatment of motivational problems. A great deal can be said for such an objection but the fact remains that the term drive is firmly entrenched in the current literature of psychology and has been given a variety of different meanings. The present paper seeks to reconcile these meanings and to give, from the point of view of the experimentalist, a definite form to the concept which is fairly consistent with much of the present practice. *Drive* as defined here is obviously placed within the organism where it properly belongs. It is, moreover, a convenient term for the specification of those extensive organic changes resulting from deprivation which have such important effects on behavior.

Characteristics of Driven Behavior

The following list of characteristics of behavior under conditions of drive has been drawn from various sources. It is difficult to reduce the different phenomena to strict comparability. Some are relatively temporary, paralleling, for example, the hunger-contractions of the stomach while others are cumulative as deprivation progresses. The items are neither coördinate nor independent but taken as a whole they

[1] In the case of the hunger-drive it is possible that such definition could be roughly approximated at the present time.

show something of the complexity of the influences which drive has on behavior. Studies of the hunger-drive have doubtless had undue importance in shaping this list because the hunger-drive has been most thoroughly investigated. Nevertheless the available evidence suggests that the different drives show rather similar effects on behavior.

1. *Activity*

A drive is primarily manifested by activity. This relationship is so evident as to be almost tautological. Whenever an organism is active it may be assumed that a drive is present, that the organism is affected by some kind of deprivation. Fundamentally all of the drives show effects on the activity although the relationship of deprivation to activity is rarely simple and linear. It is of course true that a drive may produce 'directed' activity, that is to say, it may produce activity which seems to be primarily aimed at getting food, water, or whatever the appropriate goal-object may be. For purposes of the present argument, however, it is assumed that directed or goal-seeking activity can appear only as a secondary or acquired resultant of repeated experiences. The characteristics of the activity furnish only the barest of clues as to what drive is operating. A wide variety of experiments (**6**, **10**, **15**, **16**) on animals and persons, infants and adults, can be cited as indicating that the activity resulting from drive is characteristically random in the sense that all kinds of activity appear. Under conditions of drive an organism is active, however the activity be measured, whether in terms of the number of revolutions of a revolving cage, in terms of the number of crossings in the obstruction apparatus, or in terms of the number of arm-movements.

2. *Rhythm*

Typically the drive is rhythmic in its effects. (One possible exception to this rule appears to be the sexual drive of the male where no clear cut and regular rhythms have been demonstrated.) Ordinarily the organic complex constituting the drive is either most acute in some particular organ or most

effective in stimulating receptors in that organ. In each drive there are periodic changes in this organ correlated with rhythmic behavioral effects. These rhythmic changes may persist during relatively long periods of deprivation (**10**)—a fact too often neglected by experimenters measuring strength of drive.

3. *Facilitation*

All the available evidence points to a widespread and general facilitation of nervous processes underlying the activity which is characteristic of driven behavior. Such facilitation certainly occurs in the case of the hunger-drive, and, while less is known of the other drives, the available evidence points in the same direction. Numerous examples of this facilitation can be given. The knee-jerk and other responses in man are more strongly elicited during hunger (**4, 15**). The strength of grip is increased (**15**). Chickens peck at grains of food with a vigor dependent upon the extent of deprivation (**1, 3**). With hunger-contractions the heart-beat increases, and the pancreatic, salivary, and other secretions also increase (**4**). There is some indication that even scores on intelligence tests are improved during hunger (**15**).

The facilitation is extensive and general and it is not limited, as one might suppose, to the 'appropriate' reflexes such as sucking or pecking in hunger. The facilitation appears to furnish the basis for some of the increased activity under drive. It may be objected that the distinction made here between facilitation and activity as characteristics of driven behavior is an arbitrary one. However, as indicated above, the facilitation is much more extensive than any existing measures of activity would lead one to expect.

4. *Inhibition*

Physiological experimentation indicates that facilitation of some processes generally means the inhibition of other processes (**12**). The evidence for inhibition as a characteristic of driven behavior is not very extensive. It is true that as soon as behavior is modified by learning, something like inhibition appears in motivated behavior. The 'direction' of motivated

behavior seems to imply a conditioned inhibition. As a result of learning the appropriate food responses are selected out of the organism's whole repertoire and come to be evoked more readily and more often during hunger.

Recent results (**2**) on the food reactions of rats have a more direct bearing on inhibition as a characteristic of driven behavior. With increase in the length of the fast there is a decrease in the rate of eating and also in the amount of food eaten. Preliminary evidence on another experiment (**3**) indicates that the accuracy with which a chicken pecks at grains of food is greatly decreased by strong hunger-drive. This latter evidence might be interpreted in terms of either facilitation or inhibition.

5. *Sensitivity*

The relationship of sensitivity to drive has not been investigated thoroughly. There are, however, a number of scattered and more or less incidental observations which indicate that drive is associated with an increase in sensitivity. Persons undergoing fasts have reported that they became increasingly sensitive to sounds or odors but these statements have not been verified experimentally. In one intensive study (**7**) a large increase in visual acuity has been reported. With pigeons it requires a much louder sound to cause the lowering of the feathers or opening of the eyes just after the animals have been fed than when they are hungry (**11**). It is also reported that the sucking reflex in babies is more easily elicited during hunger (**17**)!

6. *Variability*

Experimenters usually report that weakly motivated animals are more variable and less dependable in their behavior. Exact evidence on this point is scarce. It has been shown (**5**) that strong hunger-drive under certain conditions tends to reduce the variability of performance. Recent work (**2**) on rats indicates that the times for eating individual pellets of food become relatively less variable as the length of the fast is increased.

The fact that drive is characterized by activity and general facilitation seems to imply that it should also be associated with variability. There is probably some such fundamental association, although it may not appear, for learning can so affect behavior that, in a particular situation, the strongest drive may cause the least variable activity. 'Persistence of response' which is sometimes taken as a characteristic of motivated behavior (9) is obviously the reverse of variability. The organism is persistent when its behavior is not particularly variable and this condition is a result of learning rather than an inherent characteristic of driven behavior.

7. *Modifiability*

Ease of modification [cf. Tolman's 'docility' (14)] is often accepted as a characteristic of driven behavior, and numerous experiments have been supposed to demonstrate this relationship. Unfortunately most of the supposedly relevant experiments have dealt with rewards rather than drives and therefore are not entirely convincing. The crucial question would be to ask whether drive facilitates learning only secondarily by making a reward more acceptable and effective, or whether drive is fundamentally a state of greater modifiability independently of the acceptability of the reward. This latter possibility is more or less reasonable but has not been demonstrated experimentally.

Measurement of Strength of Drive

The condition of drive has, as indicated above, many and varied effects on behavior. These effects appear as complex and shifting characteristics and changes of behavior. By the very nature of the facts to be subsumed a drive is conceived as something which varies in strength, and again naturally enough there are a large number of experiments purporting to measure the relative or absolute strengths of different drives. Certain inadequacies in such studies of strength of drive become increasingly evident as the numerous behavioral effects of drive are examined. Most experimenters in attempting to measure the strength of drive have failed to recognize the

spread of the behavioral effects. They have ordinarily selected for investigation some single aspect of behavior and have taken the intensity, duration, or frequency of this narrow bit of behavior as the measure of strength of drive. For example, the number of times a rat crosses a charged grid has been used as measure of its hunger-drive (**16**). The extensive changes under conditions of drive which we have examined here indicate the arbitrary nature of any such selection of a single measure for strength of drive.[2] Moreover, despite the fact that the underlying conditions constituting the drive become more and more acute with fasting, this particular measure shows first an increase for a few days and then a decrease.

The case of the sexual drive of male rats as measured by this same obstruction method (**16**) makes the arbitrary nature of the method of measurement even more evident. With increasing periods of sexual deprivation the strength of the drive appears gradually to decrease if it is measured by the tendency to cross a charged grill to the female. However, it is extremely probable that, if the sexual drive of male rats were measured in terms of the frequency of perversions, it would appear to increase in strength with the time of deprivation. The obstruction method furnishes interesting information about the changing patterns of the behavioral effects of drive, but it can not be exalted into a general measure of the strength of drive.

A different treatment of strength of drive has been given by Skinner (**13**). He identifies strength of drive with reflex strength as he can readily do because of the relationship of facilitation to drive. Nor is it necessary to take issue with this view except to note that strength of drive can not be adequately defined in terms of changes in a single reflex or group of reflexes only. By defining the strength of the hunger-drive in terms of the eating reflexes alone, Skinner neglects the widespread nature of the facilitation underlying drive.

[2] Leuba (**8**) has also criticized these studies and has shown that a slight change in the method of measurement will give entirely different results as to the relative strengths of the different drives.

The effects of drive are so extensive and general that there is no single adequate measure of its strength in terms of behavior. There is good reason, therefore, to limit, for the present, the specification of strength of drive to terms of deprivation. This state of affairs may not prove satisfactory to the theorist but should tend to clarify matters for the experimentalist.

Summary

1. Motivational experiments deal primarily with deprivation and seek relationships between deprivation and behavior.

2. Drive is an entity conceived as intervening between deprivation and behavior. It is most usefully defined as the complex of organic conditions resulting from and controlled by deprivation.

3. Activity, rhythm, facilitation, inhibition, sensitivity, variability, and modifiability, are among the characteristics of driven behavior.

4. The effects of drive on behavior are too extensive to permit the use of any single behavioral measure of strength of drive.

REFERENCES

1. Bayer, E., Beiträge zur Zweikomponententheorie des Hungers, *Zsch. f. Psychol.*, 1929, **112**, 1–54.
2. Bousfield, W. A., and Elliott, M. H., The effect of fasting on the eating behavior of rats, *J. Genet. Psychol.*, 1934, **45**, 227–237.
3. Bousfield, W. A., and Spear, E., The influence of hunger on the pecking responses of chickens, *Amer. J. Psychol.*, in press.
4. Carlson, A. J., The control of hunger in health and disease: Chicago, Univ. of Chicago Press, 1916.
5. Elliott, M. H., The effect of hunger on variability of performance, *Amer. J. Psychol.*, 1934, **46**, 107–112.
6. Irwin, O. C., The amount and nature of activities of newborn infants under constant external stimulating conditions during the first ten days of life, *Genet. Psychol. Monog.*, 1930, **8**, 1–92.
7. Langfeld, H. S., The psychophysiology of a prolonged fast, *Psychol. Monog.*, 1914, **16**, no. 6.
8. Leuba, C. J., Some comments on the first reports of the Columbia study of animal drives, *J. Comp. Psychol.*, 1931, **11**, 275–279.
9. McDougall, W., Outline of psychology: New York, Scribner's, 1923.
10. Richter, C. P., A behavioristic study of the activity of the rat, *Comp. Psychol. Monog.*, 1922, **1**, no. 2.

11. Rogers, F. T., Contributions to the physiology of the stomach. XXXIX. The hunger mechanism of the pigeon and its relation to the central nervous system, *Amer. J. Physiol.*, 1916, **41**, 555–570.
12. Sherrington, C. S., The integrative action of the nervous system: New Haven, Yale Univ. Press, 1906.
13. Skinner, B. F., Drive and reflex strength: I and II, *J. Gen. Psychol.*, 1932, **6**, 22–48.
14. Tolman, E. C., Purposive behavior in animals and men: New York, Century, 1932.
15. Wada, T., An experimental study of hunger in relation to activity, *Arch. Psychol.*, 1922, **8**, no. 57.
16. Warden, C. J., Animal motivation studies: The albino rat: New York, Columbia Univ. Press, 1931.
17. Watson, J. B., Psychology from the standpoint of a behaviorist: Philadelphia, Lippincott, 2nd ed., 1924.

[MS. received October 8, 1934]

NOTES

A NOTE ON THE RECALL OF THE PLEASANT VS. THE UNPLEASANT

The relationship between feeling and memory was so well analyzed by Meltzer (1) in 1930 that any new study will add little to the evaluation of the experiments prior to that date. Since the printing of Meltzer's article, there have appeared at least nine additional studies in the field of affective recall.

It will be remembered that Meltzer seriously questioned the Freudian interpretation of forgetting. Most of the twenty-six studies which he reported were full of shortcomings and revealed little agreement as to the influence of feeling upon memory. Sixteen suggested that 'pleasant feeling tone facilitates memory,' while ten tended to disprove such a conclusion. It is perhaps significant that the first thirteen studies appearing between 1898 and 1919 tended to support the Freudian hypothesis by a vote of ten to three while the thirteen studies appearing in the following decade reversed this stand by a vote of seven to six. These conclusions are not significant in view of the insufficient number of subjects represented by four, faulty and unreliable methods found in eight, unwarranted assumptions made in four, unlifelike conditions in five, and unwarranted interpretations in three. The two most satisfactory studies reported, those of Wohlgemuth (2) and Gordon (3) revealed each a difference of less than one per cent between the tendency to remember the pleasant and the unpleasant.

A review of the nine experiments since 1929 evidence the influence of Meltzer's suggestions and criticisms. The previous studies for the most part either failed to determine the proportion of original pleasant and unpleasant items or else provided an arbitrary classification of them. None of the recent studies is faulty in these respects. In all cases the proportion of pleasant and unpleasant has been determined at the time of experience or immediately following by each individual subject. Since the method used in these later experiments was standardized, the results readily lend themselves to a comparative analysis. The agreement in conclusions is quite in contrast to the disagreements which appeared prior to such stand-

ardization. Whether the items of recall are school grades (4), previous experiences (5, 6, 7) nonsense syllables paired with odors (8), words (9), or poems, news events and diary records (10),—in all cases the recall of the pleasant is more efficient than that of the unpleasant. In one case the advantage for the pleasant reaches as high as 45 per cent.[1] Cason (11), who found no significant difference in the recall of the pleasant and unpleasant, provides the only exception to the conclusions stated above. While he discovered no general differences in affective recall, his study suggests that those with optimistic temperament tend to recall more of the pleasant, and those with pessimistic temperament recall more of the unpleasant. While the recall of the pleasant is consistently greater than of the unpleasant, it should be recognized that in many of the experiments this advantage was slight.

REFERENCES

1. Meltzer, H., Present status of experimental studies on relationship of feeling to memory, Psychol. Rev., 1930, 37, pp. 124–39.
2. Wohlgemuth, A., The influence of feeling on memory, *Brit. J. Psychol.*, 1922–23, 13, pp. 405–29.
3. Gordon, K., Recollection of pleasant and unpleasant odors, *J. Exper. Psychol.*, 1925, 8, pp. 225–39.
4. Koch, H. L., The influence of some affective factors for recall, *J. Gen. Psychol.*, 1930, 4, pp. 171–190.
5. Meltzer, H., The forgetting of pleasant and unpleasant experiences in relation to intelligence and achievement, *J. Soc. Psychol.*, 1930, 2, pp. 217–27.
6. Stagner, R., Redintegration of pleasant and unpleasant experiences, *Amer. J. Psychol.*, 1931, 43, pp. 463–8.
7. Jersild, A., Memory for the pleasant as compared with the unpleasant, *J. Exper. Psychol.*, 1931, 14, pp. 284–288.
8. Frank, J. P. and Ludvigh, E. J., The retractive effect of pleasant and unpleasant odors on learning, *Amer. J. Psychol.*, 1931, 43, pp. 102–8.
9. Stagner, R., Factors influencing the memory value of words in a series, *J. Exper. Psychol.*, 1933, 16, pp. 129–137.
10. Thompson, R. N., Experimental study of memory as influenced by feeling tone, *J. Exper. Psychol.*, 1930, 13, pp. 462–8.
11. Cason, H., The learning and retention of pleasant and unpleasant activities, *Arch. Psychol.*, 1932, 134, pp. 96.

Elon H. Moore

Oregon State College

[MS. received August 10, 1934]

[1] Lest the reader interpret the 45 per cent appearing here as a misuse of the 45 per cent given by Jersild in his conclusions, it should be explained that pleasant/unpleasant (45/31) yields 145 per cent (7).

A NOTE ON "THE VECTORS OF MIND"[1]

The elaboration of statistical techniques for factor analysis though originated recently by psychologists, has apparently out run the psychological developments of the problem. The meticulosity with which factor values are computed and then the laxity with which the factors (or groups of abilities or traits) are identified is easily observed. An illustration may be cited from "The Vectors of Mind" in which is described the results obtained from analyzing certain data supplied by Mr. Alexander. Regarding the factors Professor Thurstone writes "There seems to be little doubt in naming one of these constellations *verbal ability*, and we shall name the other one tentatively as *manipulation*, since that is common to all of the performance tests."[2] Concerning the same data and treatment, Mr. Alexander writes "we decide with excellent grounds to call (the first) the verbal factor" . . . and of the other "In view of a similar examination of the tests we call it the practical factor."[3] Subjectivity is apparent, for "practical" and "manipulative" seem sufficiently different to lead subsequent work in diverse directions.

In presenting the results of a study "to ascertain whether radicalism is a common factor in people's attitudes on various disputed issues,"[4] it is to be noted that "the factorial analysis reveals a conspicuous common factor that we recognize as radicalism."[5] "Recognize" may almost imply prejudice on the part of the investigator even though this caution is added: "In naming the common factor which is the most prominent in these attitude scales there may be some question as to whether it should be called conservatism or religion. . . . It may also be that what we have called radicalism-conservatism is a factor which is nearly the same as the religious factor."[6] "Liberal" is used in some places as a synonym for "radical."

Since the common factor that was found can be used to account for a considerable portion of the variance of the attitude scores and since a high positive loading of the common factor characterizes a

[1] The author thanks Professors L. L. Thurstone and H. E. Burtt for their suggestions and criticisms.

[2] L. L. Thurstone, The vectors of mind, PSYCHOL. REV., 1934, 41, 1–32.

[3] W. P. Alexander, Research in guidance: A theoretical basis, *Occupations*, 1934, 12, No. 8, 75–91.

[4] L. L. Thurstone, *op. cit.*

[5] *Ibid.*

[6] L. L. Thurstone, The vectors of mind, PSYCHOL. REV., 1934, 41, 1–32.

very "radical" attitude, and a large negative loading a "conservative" attitude, then these attitudes seem to be arranged in a scale from that with large negative loading (conservative) up to that with a high positive loading (radical). If, now, these attitudes be scaled in another experiment with respect to radicalism (or whatever the factor is), the scale values should correspond to and check the factor loadings—provided, of course, the factor has been accurately identified and provided the factor is a psychological entity, not merely a statistical artifact.

Following this reasoning an attempt was made to check the factor loadings as well as to find some additional evidence as to whether religiousness or radicalism is the more accurate name for this common factor. A group of 111 students ranked the possession of favorable attitudes towards these issues on the basis of radicalism, *i.e.* each student was given a blank containing the eleven issues (see table) and these directions: "People who have favorable attitudes towards certain topics are called radicals, and it is of interest to find how radical certain topics are considered. Mark 1 before the term which it would be most radical to have a favorable attitude toward. Mark 2 before the term which it would be the next most radical to favor; 3 before the next, and so on to the least radical. Be sure you have ranked all eleven items as to radicalism." To a second group of 113 these items but with the directions changed to "religious" instead of "radical" were given. These ranks were averaged and are shown in the table. If these averages (assumed to represent

Attitude	Factor Loading (Estimated)	Average rank for:	
		Radicalism	Religiousness
Church	0.86	7.4	1.9
God	.85	8.4	1.2
Sunday observance	.75	6.8	3.2
Patriotism	.64	7.4	4.8
Prohibition	.39	5.4	5.3
War	.16	3.3	9.4
Germans	.09	6.9	7.1
Birth Control	−.46	5.2	7.5
Communism	−.67	3.1	8.8
Evolution	−.73	5.7	8.6
Divorce	−.75	5.4	8.0

scale values) be compared with the factor loadings, it seems that religiousness is more nearly correct than radicalism as the name for the common factor. This may be due to the large role played by

religious issues in the study, and is supported by the high relationship between the common factor and the attitudes toward Church and God. This similarity between the two series is to be noted.

In conclusion, factors which are easily isolated can not always be readily identified, yet identification probably is necessary for any additional study of the factor. In some case it appears that after a judicious guess has been made, the use of techniques such as scaling may be of value. High loadings (correlations) between the factor and a variable can be examined for cues which will be helpful in making the guess.

HERMAN A. COPELAND

THE OHIO STATE UNIVERSITY

[MS. received August 14, 1934]

PSYCHOLOGICAL REVIEW PUBLICATIONS

Original contributions and discussions intended for the Psychological Review should be addressed to

Professor Herbert S. Langfeld, Editor PSYCHOLOGICAL REVIEW,
Princeton University, Princeton, N. J.

Original contributions and discussions intended for the Journal of Experimental Psychology should be addressed to

Professor Samuel W. Fernberger, Editor JOURNAL OF EXPERIMENTAL PSYCHOLOGY,
University of Pennsylvania, Philadelphia, Pa.

Contributions intended for the Psychological Monographs should be addressed to

Professor Joseph Peterson, Editor PSYCHOLOGICAL MONOGRAPHS,
George Peabody College for Teachers, Nashville, Tenn.

Reviews of books and articles intended for the Psychological Bulletin, announcements and notes of current interest, and *books offered for review* should be sent to

Professor John A. McGeoch, Editor PSYCHOLOGICAL BULLETIN,
University of Missouri, Columbia, Mo.

Titles and reprints intended for the Psychological Index should be sent to

Professor Walter S. Hunter, Editor PSYCHOLOGICAL INDEX,
Clark University, Worcester, Mass.

All business communications should be addressed to

Psychological Review Company, Princeton, New Jersey

THE PSYCHOLOGICAL REVIEW

is indexed in the

International Index to Periodicals

to be found in most public and
college libraries

AMERICAN PSYCHOLOGICAL PERIODICALS

American Journal of Psychology—Ithaca, N. Y.; Cornell University.
Subscription $6.50. 624 pages annually. Edited by M. F. Washburn, Madison Bentley, K. M. Dallenbach, and E. G. Boring.
Quarterly. General and experimental psychology. Founded 1887.

Journal of Genetic Psychology—Worcester, Mass.; Clark University Press.
Subscription $14.00 per yr.; $7.00 per vol. 1,000 pages ann. (2 vols.). Edited by Carl Murchison.
Quarterly. Child behavior, animal behavior, comparative psychology. Founded 1891.

Psychological Review—Princeton, N. J.; Psychological Review Company.
Subscription $5.50. 540 pages annually. Edited by Herbert S. Langfeld.
Bi-monthly. General psychology. Founded 1894.

Psychological Monographs—Princeton, N. J.; Psychological Review Company.
Subscription $6.00 per vol. 500 pages. Edited by Joseph Peterson.
Without fixed dates, each issue one or more researches. Founded 1895.

Psychological Index—Princeton, N. J.; Psychological Review Company.
Subscription $4.00. 400–500 pages. Edited by Walter S. Hunter and R. R. Willoughby.
An annual bibliography of psychological literature. Founded 1895.

Psychological Bulletin—Princeton, N. J.; Psychological Review Company.
Subscription $6.00. 720 pages annually. Edited by John A. McGeoch.
Monthly (10 numbers). Psychological literature. Founded 1904.

Archives of Psychology—New York, N. Y.; Columbia University.
Subscription $6.00. 500 pages per volume. Edited by R. S. Woodworth.
Without fixed dates, each number a single experimental study. Founded 1906.

Journal of Abnormal and Social Psychology—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $5.00. 448 pages annually. Edited by Henry T. Moore.
Quarterly. Abnormal and social. Founded 1906.

Psychological Clinic—Philadelphia, Pa.; Psychological Clinic Press.
Subscription $3.00. 288 pages. Edited by Lightner Witmer.
Without fixed dates (Quarterly). Orthogenics, psychology, hygiene. Founded 1907.

Journal of Educational Psychology—Baltimore: Warwick & York.
Subscription $6.00. 720 pages. Monthly except June to August.
Edited by J. W. Dunlap, P. M Symonds and H. E. Jones. Founded 1910.

Psychoanalytic Review—Washington, D. C.; 3617 10th St., N. W.
Subscription $6.00. 500 pages annually. Edited by W. A. White and S. E. Jelliffe.
Quarterly. Psychoanalysis. Founded 1913.

Journal of Experimental Psychology—Princeton, N. J.; Psychological Review Company.
Subscription $7.00. 900 pages annually. Edited by Samuel W. Fernberger.
Bi-monthly. Experimental psychology. Founded 1916.

Journal of Applied Psychology—Indianapolis; C. E. Pauley & Co.
Subscription $5.00. 600 pages annually. Edited by James P. Porter, Ohio University, Athens, Ohio. Bi-monthly. Founded 1917.

Journal of Comparative Psychology—Baltimore, Md.; Williams & Wilkins Company.
Subscription $5.00 per volume of 450 pages. Ed. by Knight Dunlap and Robert M. Yerkes. Two volumes a year. Founded 1921.

Comparative Psychology Monographs—Baltimore, Md.; The Johns Hopkins Press.
Subscription $5.00. 400 pages per volume. Knight Dunlap, Managing Editor.
Published without fixed dates, each number a single research. Founded 1922.

Genetic Psychology Monographs—Worcester, Mass.; Clark University Press.
Subscription $7.00 per vol. One volume per year. Edited by Carl Murchison.
Bi-monthly. Each number one complete research. Child behavior, animal behavior, and comparative psychology. Founded 1925.

Psychological Abstracts—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $6.00. 700 pages ann. Edited by Walter S. Hunter and R. R. Willoughby.
Monthly. Abstracts of psychological literature. Founded 1927.

Journal of General Psychology—Worcester, Mass.; Clark University Press.
Subscription $14.00 per yr.; $7.00 per vol. 1,000 pages ann. (2 vols.). Edited by Carl Murchison.
Quarterly. Experimental, theoretical, clinical, historical psychology. Founded 1927.

Journal of Social Psychology—Worcester, Mass.; Clark University Press.
Subscription $7.00. 500 pages annually. Ed. by John Dewey and Carl Murchison.
Quarterly. Political, racial, and differential psychology. Founded 1929.